CUL-H02711-55-P014903

55

AMERICAN JOURNAL OF MATHEMATICS

FOUNDED BY THE JOHNS HOPKINS UNIVERSITY

EDITED BY

E. T. BELL
CALIFORNIA INSTITUTE OF TECHNOLOGY

ABRAHAM COHEN
THE JOHNS HOPKINS UNIVERSITY

T. H. HILDEBRANDT
UNIVERSITY OF MICHIGAN

F. D. MURNAGHAN
THE JOHNS HOPKINS UNIVERSITY

J. F. RITT
COLUMBIA UNIVERSITY

WITH THE COÖPERATION OF

MARSTON MORSE
E. P. LANE
ALONZO CHURCH
L. R. FORD
OSCAR ZARISKI
G. C. EVANS
AUREL WINTNER
GABRIEL SZEGÖ
R. L. WILDER
R. D. JAMES
OYSTEIN ORE
H. P. ROBERTSON
M. H. STONE
T. Y. THOMAS
G. T. WHYBURN

PUBLISHED UNDER THE JOINT AUSPICES OF

THE JOHNS HOPKINS UNIVERSITY
AND
THE AMERICAN MATHEMATICAL SOCIETY

CALCUTTA UNIVERSITY CENTRAL LIBRARY

510·5
Ao2

VOLUME LX

1938

THE JOHNS HOPKINS PRESS
BALTIMORE, MARYLAND
U. S. A.

CUTTA
CENTRAL LIBRARY.
UNIVERSITY.

INDEX

PAGE

Adkisson, Virgil W. See MacLane and Adkisson, page 611.

Bateman, H. A partial differential equation associated with Poisson's work on the theory of sound, ... 293

Bateman, H, and Rice, S. O. Integrals involving Legendre functions, 297

Boas, R. P. Jr. and Smithies, F. On the characterization of a distribution function by its Fourier transform, ... 523

Brown, E. W. The equations of motion of the Moon, ... 785

Busemann, Herbert. On translations in general plane geometries, . 227

Cairns, Stewart S. Normal coördinates for extremals transversal to a manifold, ... 423

Camp, Chester C. On multiparameter expansions associated with a differential system and auxiliary conditions at several points in each variable, ... 447

Chazy, Jean. Sur les avances du noeud et du périhélie d'une planète sous l'action d'une anneau circulaire, ... 793

Dresher, Melvin and Ore, Oystein. Theory of multigroups, . . 705

Farrell, O. J. On the representation of bounded analytic functions by sequences of polynomials, ... 573

Fogels, E. The general rational solution of the equation $ax^2 - by^2 = z^3$, 734

Gottlieb, Morris J. Concerning some polynomials orthogonal on a finite or enumerable set of points, ... 453

Green, Louis. Systems of quadrics associated with a point of a surface, 649

Griffiths, L. W. On hypergroups, multigroups, and product systems, 345

Hamilton, H. J. and Hill, J. D. On strong summability, . . . 588

Hartman, Philip. On a class of arithmetical Fourier series, . . 66

Hartman, Philip and Kershner, Richard. Integral forms and variational orthogonality, ... 205

Hartman, Philip and Kershner, Richard. On the Fourier-Stieltjes transform of a singular function, ... 459

Hartman, Philip, van Kampen, E. R., and Wintner, Aurel. On the distribution functions of almost periodic functions, . . . 491

Hedlund, Gustav A. See Morse and Hedlund, page 815.

ill, J. D. See Hamilton and Hill, page 588.

ölder, Ernst. Die symmetrischen periodischen Bahnen des restringierten Dreikörperproblems in der Nachbarschaft eines kritischen Keplerkreises, ... 801

ham, M. H., and Wolf, M. C. Convergence of a sequence of linear transformations, ... 107

PAGE

Jackson, Dunham. Problems of closest approximation on a two-dimensional region, ... 436

James, R. D. The distribution of integers represented by quadratic forms, ... 737

Kershner, Richard. Determination of a van der Corput absolute constant, ... 549

Kershner, Richard. See Hartman and Kershner, page 205.

Kershner, Richard. See Hartman and Kershner, page 459.

Kline, Morris. Homomorphism of rings and fields of point sets, . 139

Lane, E. P., and MacQueen, M. L. Surfaces whose asymptotic curves are twisted cubics, ... 337

Lane, E. P., and MacQueen, M. L. Asymptotic curves on a surface, . 640

Lewis, Daniel C., Jr. Invariant manifolds near an invariant point of unstable type, ... 577

Lewy, Hans. A property of spherical harmonics, ... 555

MacLane, Saunders, and Adkisson, Virgil W. Fixed points and the extension of the homeomorphisms of a planar graph, . . . 611

MacQueen, M. L. See Lane and MacQueen, page 337.

MacQueen, M. L. See Lane and MacQueen, page 640.

McCoy, Neal H. Subrings of direct sums, ... 374

McEwen, W. H. A note on an extension of Bernstein's theorem, . . 309

Mehlenbacher, Lyle E. The interrelations of the fundamental solutions of the hypergeometric equation; logarithmic case, . . 120

Miller, G. A. Groups whose commutator subgroups are of order two, 101

Modesitt, Virginia. Some singular properties of conformal transformations between Riemannian spaces, ... 325

Morse, Marston, and Hedlund, Gustav A. Symbolic dynamics, . 815

Murnaghan, F. D. The analysis of the direct product of irreducible representations of the symmetric groups, ... 44

Murnaghan, F. D. The analysis of the Kronecker product of irreducible representations of the symmetric group, ... 761

Newsom, C. V. On the character of certain entire functions in distant portions of the plane, ... 561

Ore, Oystein. See Dresher and Ore, page 705.

Pitt, H. R. An extension of Wiener's general Tauberian theorem, . 532

Rademacher, Hans. The Fourier coefficients of the modular invariant $J(\tau)$, ... 50

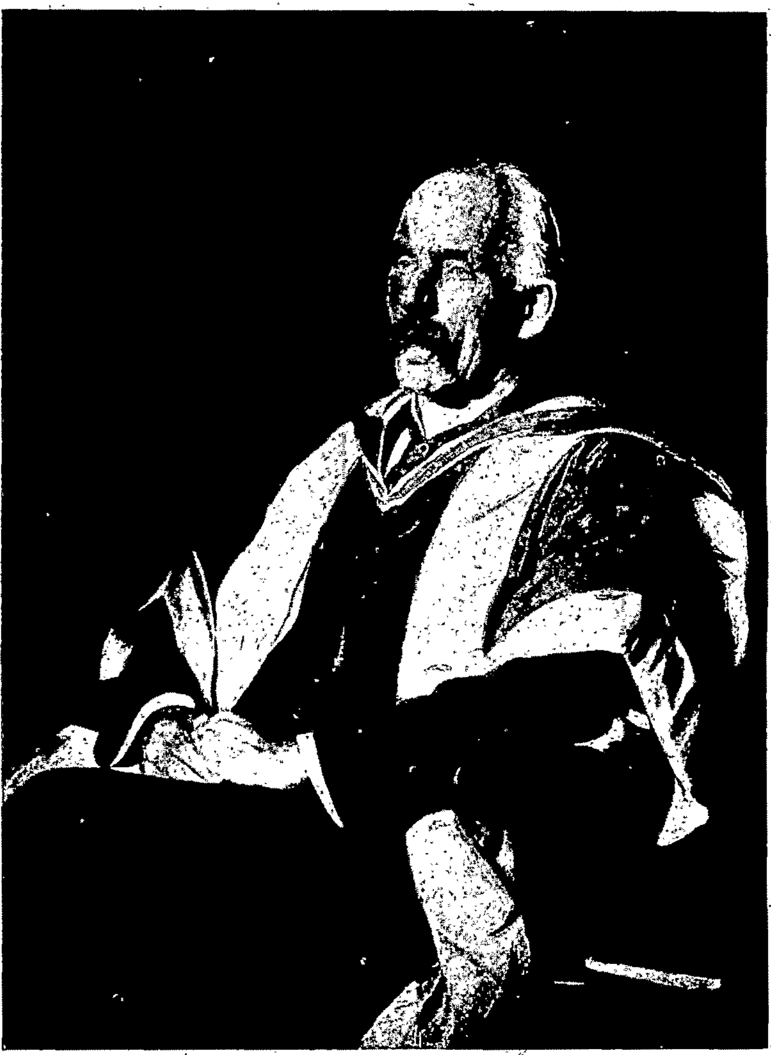
Reid, William T. An integro-differential boundary value problem, . 2

Rice, S. O. See Bateman and Rice, page 297.

Ritt, J. F. On certain points in the theory of algebraic differential equations, ...

Ritt, J. F. Systems of differential equations. I. Theory of ideals, .

PAGE
Robinson, G. de B. On representations of the symmetric group, 745
Schilling, O. F. G. The structure of local class field theory, 75
Schilling, O. F. G. A generalization of local class field theory, 667
Schilling, O. F. G., and Zariski, O. On the linearity of pencils of curves on algebraic surfaces, 320
Smithies, F. See Boas and Smithies, page 523.
Strömgren, Elis. Periodic orbits in the restricted problem of three bodies in their relation to Hill's work on the motion of the Moon, 867
Taylor, A. E. Biharmonic functions in abstract spaces, 416
Thrall, Robert M. Metabelian groups and trilinear forms, 383
Toeplitz, Otto. Zur Theorie der Dirichletschen Reihen, 880
van Kampen, E. R. The theorems of Gauss-Bonnet and Stokes, 129
van Kampen, E. R. Invariants derived from looping coefficients, 595
van Kampen, E. R. See Hartman, van Kampen and Wintner, page 491.
Weyl, Hermann. Mean Motion, 889
Whyburn, G. T. Interior transformations on surfaces, 477
Wiener, Norbert. The homogeneous chaos, 897
Wiener, Norbert and Wintner, Aurel. Fourier-Stieltjes transforms and singular infinite convolutions, 513
Williamson, John. Matrices normal with respect to an Hermitian matrix, 355
Wintner, Aurel. Liouville systems and almost periodic functions, 463
Wintner, Aurel. Galilei group and law of gravitation, 473
Wintner, Aurel. On Hill's periodic lunar orbit, 937
Wintner, Aurel. See Hartman, van Kampen and Wintner, page 491.
Wintner, Aurel. See Wiener and Wintner, page 513.
Wolf, M. C. See Ingraham and Wolf, page 107.
Zariski, Oscar. Polynomial ideals defined by infinitely near base points, 151
Zariski, Oscar. See Schilling and Zariski, page 320.



FRANK MORLEY

1860-1937

ON CERTAIN POINTS IN THE THEORY OF ALGEBRAIC DIFFERENTIAL EQUATIONS.*

By J. F. Ritt.

The topics considered in this paper are:

I. *Forms in several unknowns.*
II. *Pairs of forms.*
III. *Essential manifolds composed of one solution.*
IV. *An approximation theorem.*
V. *Essential irreducible manifolds in the manifold of a form.*
VI. *Equations in two unknowns, of the first order.*

The results, under each topic, are described at the head of the section dealing with that topic.

I. Forms in Several Unknowns.

1. Let us consider an algebraically irreducible form A in the unknowns $y_1, \cdots, y_n$. The general solution of A has $n-1$ arbitrary unknowns.[1] It is natural to inquire as to the possibilities for the number of arbitrary unknowns in the other essential irreducible manifolds in the manifold of A. This inquiry is answered by the following theorem:

Given a non-zero form[2] A *in* $y_1, \cdots, y_n$, *every essential irreducible manifold in the manifold of* A *has* $n-1$ *arbitrary unknowns.*

In other words, *every essential irreducible manifold in the manifold of* A *is the general solution of a form in* $y_1, \cdots, y_n$.

2. It will evidently suffice to prove that every solution

$$y_i = \eta_i, \qquad (i = 1, \cdots, n,) \tag{1}$$

of A is contained in an irreducible manifold, held by A, which has $n-1$ arbitrary unknowns. We wish to show that we may restrict our examination to the case of $\eta_i = 0$, $i = 1, \cdots, n$. Let the η_i in a given solution (1) be

---

* Received July 27, 1937.

[1] A set of arbitrary unknowns of a system of forms will be called a set of arbitrary unknowns for the manifold of the system.

[2] Algebraic irreducibility is not necessary.

adjoined to the underlying field.[3] Under the substitution $y_i = z_i + \eta_i$, A goes over into a form A' in $z_1, \cdots, z_n$. To prove that $z_i = 0$, $i = 1, \cdots, n$, belongs to an irreducible manifold with $n - 1$ arbitrary unknowns, held by A', will be to show that (1) is embedded as described above. Accordingly we assume in what follows that A has the solution $y_i = 0$, $i = 1, \cdots, n$, and we limit ourselves to the study of that solution.

3. Let B represent the sum of the terms of lowest degree in A considered as a polynomial in the y_{ij}. Changing the notation if necessary, we assume that B effectively involves one or more y_{1j}.

Let $v_2, \cdots, v_n$ be functions of x with a common domain of analyticity which, when substituted respectively for $y_2, \cdots, y_n$ in B, reduce B to a form C in y_1 which involves one or more y_{1j} effectively.

Representing by c an arbitrary constant, we put, in A,

$$y_i = v_i c, \qquad (i = 2, \cdots, n.) \tag{2}$$

Then A goes over into an expression D in x, c, y_1.

We shall prove the existence of a formal power series

$$\phi_1 c + \phi_2 c^{p_2} + \cdots + \phi_k c^{p_k} + \cdots, \tag{3}$$

described as in § 11 of our paper, *On the singular solutions of algebraic differential equations*,[4] which causes D to vanish identically in x and c when substituted for y_1 in D.

4. For ϕ_1, we take any function which annuls C above when substituted for y_1 in C. If D vanishes for $y_1 = \phi_1 c$, $\phi_1 c$ can be used for (3). In what follows, we assume that such vanishing does not occur.

5. In D, we put $y_1 = \phi_1 c + u_1$ where u_1 is a new unknown. Then D goes over into an expression H' in x, c, u_1. Designating by m the order of D in y_1, we arrange H' as a polynomial in $u_{10}, \cdots, u_{1m}$. We write

$$H' = a'(c) + \Sigma b'_i(c) u_{10}^{a_{0i}} \cdots u_{1m}^{a_{mi}}. \tag{4}$$

Here $a'(c)$ and the $b'(c)$ are polynomials in c with coefficients which are functions of x. The terms in Σ are those which are not free of $u_{10}, \cdots, u_{1m}$ and we understand that no b'_i is zero. We know that $a'(c)$, which equals $D(\phi_1 c)$, is not zero. As to i, it ranges from unity to some positive integer.

---

[3] If necessary, with a shrinking of the domain of meromorphicity for the field.

[4] *Annals of Mathematics*, vol. 37 (1936), p. 552. Denoted below by *S. S.*

Let σ' be the least exponent of c in a' and σ'_i the least exponent of c in $b'_i(c)$. Let

$$(5) \qquad \rho_2 = \operatorname{Max} \frac{\sigma' - \sigma'_i}{\alpha_{0i} + \cdots + \alpha_{mi}}$$

where i has the range which it has in Σ.

We shall prove that $\rho_2 > 1$.

Let r be the degree of B of § 3. Let us consider D as a polynomial in c and the y_{1j}. Then each term of D has a degree not less than r. There are terms of degree r, and their sum is obtained by making the substitution (2) in B. Because $C(\phi_1) = 0$, the sum just mentioned vanishes for $y_1 = \phi_1 c$. Thus, as $a'(c) = D(\phi_1 c)$, we have $\sigma' > r$.

Under the substitution $y_1 = \phi_1 c + u_1$, a term of D of degree s in c and the y_{1j} produces a set of terms of degree s in c and the u_{1j}. In particular, the terms of degree r in D will contribute to H' power products in c and the u_{1j} effectively involving the u_{1j} and of degree r. This means that, for certain i, we have, in (4),

$$\sigma'_i + \alpha_{0i} + \cdots + \alpha_{mi} = r.$$

For such i, we have, since $\sigma' > r$,

$$\sigma' - \sigma'_i > \alpha_{0i} + \cdots + \alpha_{mi}.$$

This, by (5), shows that $\rho_2 > 1$.

6. We now take over §§ 12-15 of *S. S.*, substituting y_1 for u_0 and the expression "solution of D of type (3)" for the expression "solution of H of type (15)." Using the argument of § 16 of *S. S.*, we have the series (3) sought for D.

7. Let us suppose now that the solution $y_i = 0$, $i = 1, \cdots, n$, of A, is not contained in a manifold with $n - 1$ arbitrary unknowns. In a decomposition of A into closed essential irreducible systems, let $\Sigma_1, \cdots, \Sigma_t$ be those systems which admit the solution $y_i = 0$, $i = 1, \cdots, n$. Let A_i be a non-zero form in Σ_i, $i = 1, \cdots, t$, involving only $y_2, \cdots, y_n$. Let $E = A_1 A_2 \cdots A_t$.

We choose v_i as in § 3 which do not annul the sum of the terms of lowest degree in E. Then E does not vanish identically in x and c for $y_i = v_i c$, $i = 2, \cdots, n$. Thus, in the decomposition of A, there is some essential irreducible system distinct from $\Sigma_1, \cdots, \Sigma_t$ whose forms all vanish for $y_i = v_i c$, $i = 2, \cdots, n$ and for y_1 as in (3).[5] Such a system must admit the solution $y_i = 0$, $i = 1, \cdots, n$. This completes the proof of the theorem stated in § 1.

---

[5] Cf. *S. S.*, § 17. Note that A vanishes for the indicated substitutions.

II. Pairs of Forms.

8. We prove the following theorem. *Let A and B be non-zero forms in $y_1, \cdots, y_n$. Let B hold A. Let A_1 be the sum of the terms of lowest degree in A considered as a polynomial in the y_{ij} and let B_1 be the corresponding sum for B. Then B_1 holds A_1.*

A similar result holds for the terms of highest degree.

9. *Remark.* The simplest case is that in which B is a linear combination of A and of the derivatives of A. One might expect that B_1 would then be a linear combination of A_1 and of its derivatives. We shall show by means of an example that this need not be so.

Let

$$A = y_1^2 + y^3; \quad B = 2y_2A - y_1\frac{dA}{dx} = 2y^3y_2 - 3y^2y_1^2$$

be forms in the unknown y. Then $A_1 = y_1^2$, $B_1 = B$. If B_1 were a linear combination of A_1 and its derivatives, y^3y_2 would be such a linear combination. If the weight of y_j is defined as j, y^3y_2 and y_1^2 have weight 2 and the derivatives of y_1^2 have weight in excess of 2. Thus, y^3y_2 would have to be simply a multiple of y_1^2. This proves our statement. From the expression of B in terms of A, one might now conjecture that *some power* of B_1 is a linear combination of A_1 and of the *first derivative of* A_1. In that case, some power of y^3y_2 would be such a linear combination. This is impossible, since y^3y_2 is not divisible by y_1. Actually, the cube of B_1 is linear in A_1 and its first two derivatives.

10. We enter into our proof. If A_1 is free of the unknowns, B_1 certainly holds A_1. In what follows, we assume that the terms of A_1 are of positive degree. Then A vanishes for $y_i = 0$, $i = 1, \cdots, n$.

We shall prove the permissibility of assuming that A_1 contains a term involving only the y_{1j}. Let $z_2, \cdots, z_n$ and $w_2, \cdots, w_n$ be new unknowns. Let y_i, for $i > 1$, be replaced in A_1 by $z_i + w_i$. Then A_1 goes over into a form C in y_1, the z_i and w_i. C contains terms free of the z_{ij}; the sum D of such terms in C is found by substituting w_i for y_i in A_1 for $i > 1$. Let t_2 be an integer which exceeds the order of D in y_1. On putting $w_2 = y_{1t_2}$ in D, we convert D into a non-zero form D_1 in $y_1, w_3, \cdots, w_n$. We now replace w_3 in D_1 by y_{1t_3}, where t_3 exceeds the order of D_1 in y_1. Continuing, we find a substitution

(6) $$y_i = z_i + y_{1t_i}, \qquad (i = 2, \cdots, n),$$

which converts A_1 into a form E in y_1 and the z_i, E possessing terms free of the z_{ij}. The terms of E will have the same degree as those of A_1.

The substitution (6) may be applied to A and B and will give a situation in which E takes the place of A_1. This proves the legitimacy of the assumption described above, and, in what follows, A_1 will be understood to have terms involving only the y_{1j}.

11. We now refer to § 3. Let

$$(7) \qquad y_1 = \phi_1; \; y_i = v_i, \qquad (i = 2, \cdots, n),$$

be any solution of A_1. By § 3, there is a series (3) such that A vanishes identically in x and c when y_i is replaced by $v_i c$ for $i > 1$ and y_1 is replaced by (3). Some power of B is a linear combination of A and its derivatives. Thus, B must vanish for the above replacements. This means that (7) is a solution of B_1, so that our theorem is proved.

12. The case of the terms of highest degree, mentioned in § 8, is perhaps most conveniently handled as follows. Let A_1 and B_1 represent the sums of the terms of highest degree in A and B respectively. Using unknowns $u; z_1, \cdots, z_n$, we put, in A and B,

$$(8) \qquad y_i = \frac{z_i}{u^2}, \qquad (i = 1, \cdots, n.)$$

We have then

$$A = C/u^m, \qquad A_1 = C_1/u^m$$
$$B = D/u^m, \qquad B_1 = D_1/u^m$$

with m a positive integer and C, C_1, D, D_1 forms in u and the z_i. C_1 and D_1 will be the sums of the terms of *least* degree in C and D respectively. Because B holds A, uD certainly holds C. By what precedes, uD_1 holds C_1. Because every solution of A_1 yields solutions of C_1 with $u \neq 0$, B_1 holds A_1.

III. Essential Manifolds Composed of One Solution.

13. We use the unknowns $y_1, \cdots, y_n$. We consider a system Σ composed of n forms

$$(9) \qquad y_i^{p_i} + F_i, \qquad (i = 1, \cdots, n),$$

where, for each i, p_i is a positive integer and F_i a form which either is identically zero or else is composed of terms each of which is of total degree greater than p_i in the y_{jk}, $j = 1, \cdots, n$; $k \geqq 0$.

We shall prove that *the solution of Σ given by $y_i = 0$, $i = 1, \cdots, n$, is an essential irreducible manifold in the manifold of Σ.*

14. We assume the statement to be false. Then the mentioned solution is contained in an irreducible manifold which is held by Σ but which, for some i, is not held by y_i. Fixing our ideas, let us suppose that y_1 does not hold this manifold. Then there is a value a of x at which the coefficients in the F_i are analytic such that, for every positive integer m and for every $\epsilon > 0$, Σ has a solution, analytic at a, of the type

$$y_i = \sum_{j=0}^{\infty} b_{ij}(x-a)^j, \qquad (i = 1, \cdots, n), \tag{10}$$

with

$$| b_{ij} | < \epsilon, \qquad (i = 1, \cdots, n;\ j = 0, \cdots, m), \tag{11}$$

and with $b_{10} \neq 0$.

We represent by s and c respectively two positive numbers which will be fixed later. Considering a definite solution (10), which corresponds to given m, ϵ, we put

$$z = \frac{x-a}{c}. \tag{12}$$

Then the y_i in (10) become analytic functions of z for z small. Again, we let

$$w_i(z) = c^{-s} y_i(x), \qquad (i = 1, \cdots, n), \tag{13}$$

with z and x related as in (12).

Each equation $y_i^{p_i} + F_i = 0$ goes over into an equation

$$w_i^{p_i} + \sum_{j=1}^{r} c^{\mu_j s - \nu_j} \alpha_j B_j = 0 \tag{14}$$

where the B are power products in the w_k and their derivatives with respect to z, the μ positive integers and the ν non-negative integers. Each α_j is the coefficient in F_i of the power product which produces B_j and we regard the α_j, for any c, as functions of z. It is unnecessary to express the dependence on i of Σ in (14).

If s is large, then, for every i, the $\mu_j s - \nu_j$ in (14) will all be positive. We fix s at a value large enough for this to be realized. Of course, s is independent of m, ϵ and the solution (10).

We have, by (10),

$$w_i(z) = \sum_{j=0}^{\infty} c^{-s+j} b_{ij} z^j, \qquad (i = 1, \cdots, n). \tag{15}$$

We shall now fix c. Let v be the greatest integer less than s. We choose c in such a way that the greatest of the quantities

$$c^{-s+j} | b_{ij} |, \qquad (i = 1, \cdots, n;\ j = 0, \cdots, v),$$

equals unity. This is possible because $b_{10} \neq 0$. Then, if $m \geqq v$ and if ϵ is small, c will be small.

When $1/m$ and ϵ decrease toward zero, the coefficient of z^j in (15) for a fixed j exceeding v, and for any fixed i, tends toward zero. For, if $j > v$ then $j \geqq s$.

It follows that, by decreasing $1/m$ and ϵ, we can select a sequence of solutions (10) which yields, for every i, a sequence of w_i which tends toward a polynomial which is either identically zero or else of degree v at most. The selection can be made in such a way that, for some i, the w_i converge to a polynomial distinct from zero. Fixing our ideas, we assume that the sequence of w_1 tends toward a polynomial $\gamma(z)$ distinct from zero.

We now consider (14) for $i = 1$. When c is small and the w_i are close to their polynomial limits, the expansion of Σ in (14) in powers of z will begin with a large number of small coefficients. This contradicts the fact that $\gamma^{p_1} \neq 0$, so that our result is established.

IV. An Approximation Theorem.

15. In *A. D. E.*,[6] § 74, we established a result equivalent to the following:

Let Σ be a non-trivial closed irreducible system in $y_1, \cdots, y_n$. Let F be any form which does not hold Σ. Let $\zeta_1, \cdots, \zeta_n$ be any solution of Σ, analytic in some area $\mathfrak{B}$. There exists a set of points, dense in $\mathfrak{B}$, such that, given any point a of the set, any positive integer m and any $\epsilon > 0$, Σ has a solution $\xi_1, \cdots, \xi_n$, analytic at a, which does not annul F at a, such that, for every i, each of the first $m + 1$ coefficients in the Taylor expansion of $\xi_i - \zeta_i$ at a is of modulus less than ϵ.

We are going to derive here the following stronger conclusion from the hypothesis in the above theorem. *There exists a set of points, residual*[7] *in* $\mathfrak{B}$, *such that, given any point a of the set, any two positive integers r and m and any $\epsilon > 0$, Σ has a solution $\xi_1, \cdots, \xi_n$, analytic at a, which does not annul F at a, such that, for every i, the r-th roots of $\xi_i - \zeta_i$ are analytic at a and have Taylor expansions at a in which the first $m + 1$ coefficients are all less than ϵ in modulus.*

16. *Remarks.* The residual set of points a is not offered as one of the attractions of the above result. What is noteworthy is the use of the r-th roots. Residual sets occur in the proof, and nothing is lost in using such a set in the statement of the theorem. We note at this point that, in the approximation theorem of *A. D. E.*, the dense set of points a may be replaced immediately by a residual set. In short, for the m and ϵ on page 102 of *A. D. E.*, one may use any point a of the area $\mathfrak{A}'$. Thus, the points a which may not be used for given m, ϵ are nowhere dense in $\mathfrak{A}$; the points a for which some impossible pair m, ϵ can be found form a set of the first category.

---

[6] *A. D. E.* will stand for our Colloquium Lectures.

[7] The complement in $\mathfrak{B}$ of the sum of a countable number of sets each of which is nowhere dense on $\mathfrak{B}$.

The stronger approximation theorem, which, as will be seen in § 26, is not without utility, is a first result in a program to perfect the approximation theorem of *A. D. E.* It would be natural to conjecture that $\zeta_1, \cdots, \zeta_n$ can be embedded analytically in a one-parameter family of solutions which do not annul F. That this is not so can be seen from § 90 of *S. S.*, where the singular solution $y = 0$ of

$$y_2^2 - y_1 + y = 0 \tag{16}$$

is discussed. That singular solution belongs to the general solution of (16). If (16) were satisfied, for every small h, by

$$y = \phi(x, h)$$

with ϕ analytic for x in some area and h small, and with ϕ vanishing identically in x for $h = 0$ but not vanishing identically in x and h, the first member of (16) would have a convergent y-solution. However, the only y-solution which exists is divergent for every $y \neq 0$.

Having disposed of the above conjecture, one might ask whether $\zeta_1, \cdots, \zeta_n$ cannot always be approximated uniformly in some area by solutions which do not annul F. Such a result would certainly not imply immediately the one which will be established here. For instance, as h approaches zero, $y = h^2 + hx$ approaches zero uniformly in any bounded area, but the coefficient of x in the expansion of $y^{1/2}$ at $x = 0$ does not tend toward zero with h.

17. We shall show that it will suffice to consider the case in which the ζ_i are identically zero. We adjoin the n functions ζ_i to the underlying field, limiting the domain of x if necessary. Σ becomes equivalent, for the enlarged field, to one or more essential irreducible systems. One of these systems, call it Σ', will admit a sequence of solutions which do not annul F and which approach $\zeta_1, \cdots, \zeta_n$ in the manner described in *A. D. E.*, § 74. Then $\zeta_1, \cdots, \zeta_n$ is a solution of Σ' and F does not hold Σ'. Under the substitution $y_i = z_i + \zeta_i$, Σ' goes over into an irreducible system Σ'' in the z_i; F goes over into a form G in the z_i. Then $z_i = 0$, $i = 1, \cdots, n$ will be a solution of Σ'' and G will not hold Σ''. This is enough for the justification of our statement.

18. Let P be a form in $y_1, \cdots, y_n$, given by

$$P = y_1^p(1 + A) + B \tag{17}$$

where p is a positive integer, A a form vanishing for $y_i = 0$, $i = 1, \cdots, n$, and B a form which, when written as a polynomial in y_1 and its derivatives, has no term of degree less than $p + 1$.

It is easy to prove, subjecting x and y_1 to a Painlevé transformation, that the solution $y_i = 0$, $i = 1, \cdots, n$, of P is not contained in any irreducible manifold which is held by P but not by y_1.

19. In accordance with § 17, we assume that $\zeta_i = 0$, $i = 1, \cdots, n$.

Let y_j be any one of the n unknowns. In the first stage of our proof we shall show that, given any positive integer r, there exists a dense set of values of x such that, given any a of the set and any m and ϵ, Σ has a solution $\xi_1, \cdots, \xi_n$ which does not annul F at a, the r-th roots of ξ_j being analytic at a and the first $m + 1$ coefficients in the expansions at a of

$$\xi_1, \cdots, \xi_{j-1}, \xi_j^{1/r}, \xi_{j+1}, \cdots, \xi_n$$

all having moduli less than ϵ.

If y_j holds Σ, our result is seen directly to hold. In what follows we assume that y_j does not hold Σ.

20. Let $y_1, \cdots, y_q$ be a set of arbitrary unknowns for Σ (if such a set exists). Whether or not y_j is among these arbitrary unknowns is of no importance. Let

$$A_{q+1}, \cdots, A_n \tag{18}$$

be a basic set for Σ, introducing $y_{q+1}, \cdots, y_n$. Let S_i represent the separant of A_i in (18).

No generality is lost in assuming that F of § 15 is divisible by y_j and by each S_i. In what follows, we assume such divisibility.

Let the r of § 19 be given and let an m then be selected. We make the non-restrictive assumption that, for every i, m exceeds the order of F in y_i. We consider the forms in (18) and also the first m derivatives of each of them. We secure thus a set of $(m + 1)(n - q)$ forms which we shall now regard as simple forms in the y_{ik}. The set of simple forms thus obtained will be denoted by Φ and the unknowns in Φ will be taken as those y_{ik} for which $k \leqq r_i + m$ where r_i is the highest of the orders in y_i of the forms in (18).

Let Π be the set of simple forms in the unknowns just described which vanish for all solutions of Φ which annul no S_i. It is easily proved that Π is a prime system. (Cf. *A. D. E.*, § 73.)

Because Σ has the solution $y_i = 0$, Π has a solution with every y_{ik} zero. Also Π is not held by F.

Let $p = mr$. Introducing a new unknown v, we put, in Π, $y_{j0} = v^p$. Then Π goes over into a system Λ in v and the y_{ik} distinct from y_{j0}. Let F with y_{j0} replaced by v^p be represented by C. Then C does not hold Λ.

Let $y_{j1}, \cdots, y_{jt}$ be those unknowns in Λ which arise from derivatives (proper) of y_j. We put, in Λ,

$$y_{ji} = v^{p-1} w_i, \qquad (i = 1, \cdots, t). \tag{19}$$

Then Λ becomes a system Ω in the y_{ik} with $i \neq j$, in v and the w_i. Let D be the form into which C is converted by (19). Then D does not hold Ω.

21. In a decomposition of Ω into essential prime systems, let $\Omega_1, \cdots, \Omega_s$ be those systems which are not held by D. We are going to show that there is some Ω_i each of whose forms vanishes when the unknowns are all replaced by 0.

Let this be false. Then there exists a form K, with no term free of the unknowns, such that $1 + K$ holds every Ω_i. Let g be the degree of K considered as a polynomial in the w_i. We replace each w_i in $1 + K$ by y_{ji}/v^{p-1} and multiply the resulting expression through by $v^{g(p-1)}$. We obtain a form R given by

$$R = v^{g(p-1)} + M$$

where M is a form of the following description. Every term of M which involves no y_{jk} is of the form

$$\alpha v^{g(p-1)} L$$

with α a function of x and L a power product of positive degree. Every term in M involving some y_{jk} effectively is of the type

$$\alpha v^d y_{j1}^{d_1} \cdots y_{jt}^{d_t} L \tag{20}$$

with L a power product free of the y_{jk} and with

$$d = g(p-1) - (d_1 + \cdots + d_t)(p-1).$$

Thus

$$d + p(d_1 + \cdots + d_t) > g(p-1). \tag{21}$$

In other words,

$$R = v^{g(p-1)}(1 + G) + H \tag{22}$$

where G, free of the y_{jk}, vanishes when the unknowns are all replaced by 0 and where the terms of H, all of which involve the y_{jk}, are of the type (20) with (21) holding. Furthermore, CR holds Λ.[8]

---

[8] Note that C is divisible by v.

Let ω be a primitive p-th root of unity. We replace v in R successively by $\omega^i v$, $i = 1, \cdots, p$. We obtain a set of forms $A_1, \cdots, A_p$ whose product T is a form with coefficients in $\mathfrak{F}$. T will involve v with exponents which are all multiples of p. Also, CT holds Λ and T has an expression

$$v^{g(p-1)p}(1+P)+Q \tag{23}$$

where P (whose exponents of v are all multiples of p) vanishes when the unknowns are replaced by 0 and where Q has terms of the type (20) with d divisible by p and with

$$d + p(d_1 + \cdots + d_t) > g(p-1)p. \tag{24}$$

In T, we replace v^p by y_{j0} and consider the resulting form U as a differential polynomial. Then U is of the form

$$y_j^{g(p-1)}(1+V)+W$$

where V vanishes for $y_i = 0$, $i = 1, \cdots, n$ and where the degree of each term of W in y_j and its derivatives exceeds $g(p-1)$.

Now FU holds Σ. Then U holds Σ. According to § 18, the solution $y_i = 0$, $i = 1, \cdots, n$ of U cannot be approximated by solutions of U with $y_j \neq 0$. This contradicts the fact that U holds Σ. We have thus proved the statement made at the head of the present section.

22. Fixing our ideas, let us suppose that every form in Ω_1 vanishes when the unknowns are all replaced by 0.

Considering Ω_1 and D, we refer to § 68 of *S. S.* Corresponding to each unknown in Ω_1, we find a function $\phi(x, h)$, analytic for x in some area and for h small, with $\phi(x, 0)$ equal to 0 for every x. The substitution of all of the $\phi(x, h)$ for their corresponding unknowns produces 0 for each form of Ω_1 but not for D.

Passing from Ω_1 to Λ, we have a $\phi(x, h)$ for every unknown in Λ, with each ϕ identically zero in x for $h = 0$, and with each form in Λ, but not C, reducing to zero for every x and h when the unknowns are replaced by the ϕ. Let the expansion in powers of h of the ϕ for v begin with a term in h^d with $d > 0$.[9] Then the expansion of the ϕ for any y_{jk} will, if it is not identically zero, begin with a power of h whose exponent exceeds $(p-1)d$.

We now consider Π. For the unknowns in Π we have a set of ϕ which annul the forms in Π but not F. The lowest exponent of h in the ϕ for y_{j0}

---

[9] Because D is divisible by v, the ϕ for v is not identically zero.

will be pd while the lowest exponent for any y_{jk} with $k > 0$ will exceed $(p-1)d$.

In what follows, the ϕ used will be those for Π.

23. We shall use an area $\mathfrak{A}_1$ in the plane of x which satisfies the following conditions:

(a) The coefficients in F and in the A_i in (18) are analytic in $\mathfrak{A}_1$.

(b) The ϕ are analytic for x in $\mathfrak{A}_1$ and h small.

(c) The coefficient of h^{pd} in the ϕ for y_{j0} vanishes nowhere in $\mathfrak{A}_1$.

(d) The coefficient of the lowest power of h in the result obtained by substituting the ϕ into F vanishes nowhere in $\mathfrak{A}_1$.

24. Let a be any value of x in $\mathfrak{A}_1$. We put $x = a$ in the ϕ, whereupon the ϕ go over into functions ψ of h, analytic for h small and vanishing for $h = 0$.

Every A_i in (18) vanishes when x is replaced by a and the other letters by their corresponding ψ. A similar statement holds for the first m derivatives of the A_i. Let us consider $A_{q+1}^{(m+1)}$, the $(m+1)$-st derivative of A_{q+1} in (18). In $A_{q+1}^{(m+1)}$, we replace x by a and every y_{ik} which occurs in Π by its ψ. For every $i \leqq q$, y_{i,r_i+m+1} with r_i as in § 20, may appear in $A_{q+1}^{(m+1)}$. With such letters, whether or not they figure effectively in $A_{q+1}^{(m+1)}$, we associate functions $\psi(h)$ which are identically zero and, where one of the letters figures effectively in $A_{q+1}^{(m+1)}$, we replace it in that form by 0. After these substitutions, $A_{q+1}^{(m+1)}$ becomes a linear expression in

$$(25) \qquad y_{q+1,r_{q+1}+m+1}$$

with coefficients which are functions of h. The coefficient of the letter (25) will be S_{q+1} with substitutions as above. Because F is divisible by S_{q+1} and because (d) of § 23 holds, that coefficient is not zero. When the expression obtained from $A_{q+1}^{(m+1)}$ is equated to zero, the letter (25) is determined as a function of h which is analytic, or has a pole, for $h = 0$. We treat $A_{q+2}^{(m+1)}$ similarly, substituting for the letter (25) the function of h just found. We proceed similarly with the $(m+1)$-st derivatives of the other A_i. When this step is concluded we treat in a similar way the higher derivatives of the A_i.

The net result of the total operation is as follows. We obtain a set of functions ψ of h, one ψ for every y_{ik}. The ψ are analytic, or have a pole, for $h = 0$; in particular the ψ for $k \leqq m$ are all analytic, and equal to zero, for $h = 0$. For $i \leqq q$, and for $k > m$, the ψ are zero. The ψ for y_{j0} has a least exponent of h equal to pd, while, for a y_{jk} with $0 < k \leqq m$, and with a ψ not identically zero, the least exponent exceeds $(p-1)d$. There is a $\delta > 0$ such that the ψ are all analytic for $| h | < \delta$, except, perhaps, for $h = 0$.

For h small and distinct from zero, the ψ become numbers which are derivatives in a normal solution of (18); the solution being analytic for $x = a$ and not annulling F at a. In such a solution, the y_i with $i \leqq q$ will be polynomials.

We examine y_j in these solutions. There is furnished for it, by what precedes, an expansion

$$\psi_0(h) + \psi_1(h)(x-a) + \cdots + \psi_k(h)\frac{(x-a)^k}{k!} + \cdots, \tag{26}$$

where the lowest exponent of h in ψ_0 is pd, while that in ψ_k with $k \leqq m$ exceeds $(p-1)d$.

For h small, and not zero, the r-th roots of y_j are analytic at a.

We write the series (26) in the form

$$\psi_0(h)[1 + \beta_1(h)(x-a) + \cdots]. \tag{27}$$

Then, for $k \leqq m$, the expansion about $h = 0$ of β_k is either zero or has a least exponent which exceeds $-d$.

Since $p = mr$, the least exponent of h in ψ_0 is mrd. Then the r-th roots of ψ_0 are analytic at the origin. Let $\gamma(h)$ be such an r-th root. The least exponent of h in γ is md.

One of the r-th roots of the bracket in (27) has an expansion

$$1 + \delta_1(h)(x-a) + \cdots$$

in which δ_k with $k \leqq m$ begins with a term in h of exponent greater than $-kd$.

Thus, in an r-th root of y_j as given by (26), the coefficient of $(x-a)^k$ with $k \leqq m$ will begin with a positive power of h. This is enough to prove the result stated in § 19.[10]

25. We conclude the proof of the result stated in § 15. Consider any r. Under the substitution $y_1 = v_1^r$, Σ goes over into a system Λ and F into a form F_1, in $v, y_2, \cdots, y_n$. From what goes before, it follows that some essential irreducible system Σ_1 in the decomposition of Λ contains the solution $v_1 = 0$; $y_i = 0$, $i > 1$, and is not held by F_1. We give Σ_1, with respect to y_2, the treatment accorded to Σ with respect to y_1. Continuing, we reach a system Σ_n and a form F_n in unknowns $v_1, \cdots, v_n$, such that F_n does not hold Σ_n and that $v_i = 0$, $i = 1, \cdots, n$, is a solution of Σ_n. This solution of Σ_n can be approximated at the points a of a residual set by solutions which do not annul F_n at a and which have expansions at a with as many small coefficients as one pleases. For the particular r used, the transformation $y_i = v_i^r$ gives the solutions $\xi_1, \cdots, \xi_n$ of § 15. Furthermore, the residual sets which correspond to the values 1, 2, 3, $\cdots$, of r have a residual set in common. The result of § 15 is thus completely proved.

---

[10] For m as given, any a in $\mathfrak{A}_1$ will serve for every ϵ. Thus there is a dense set of points a, in fact a residual set, which serve for all m, ϵ.

26. As an application, we consider a form

$$G = y_1^{p_1} \cdots y_n^{p_n} + A$$

where the p_i are non-negative integers whose sum is positive and where A is a form of the following description:

(a) Each term of A has a degree in the y_{ik} which exceeds $p_1 + \cdots + p_n$.

(b) Given any term L of A, and any y_j, L is either divisible by $y_j^{p_j}$ or else of degree higher than p_j in the y_{jk}.

We shall prove that the *solution* $y_i = 0$, $i = 1, \cdots, n$ *of* G *is not contained in any irreducible manifold which is held by* G *but not by* $y_1 y_2 \cdots y_n$.

This result appears not to be obtainable readily through the Painlevé transformation.

Using a positive integer r which will be fixed in a moment, we replace each y_i in G by v_i^r. If r is large enough, A will go over into a form in which each term is divisible by $v_1^{p_1 r} \cdots v_n^{p_n r}$. Let r be thus taken. Then G goes over into a form H which is the product of $v_1^{p_1 r} \cdots v_n^{p_n r}$ by a form K of the type

$$1 + B$$

where B vanishes for $v_i = 0$, $i = 1, \cdots, n$.

If our result were not true, it would follow from § 15 that K has solutions which, for suitable points a, have expansions which begin with as many arbitrarily small coefficients as one may desire. Q. E. D.

V. Essential Irreducible Manifolds in the Manifold of a Form.

27. With a view towards later applications, we shall extend here the results of *S. S.*, Part I, to forms in several unknowns.

28. Following *S. S.*, §§ 1-3, one secures the following result.

Let F and A be two forms in $y_1, \cdots, y_n$, both of class n and algebraically irreducible. Let the orders of F and A in y_n be m and $l < m$ respectively. Let A_j represent the j-th derivative of A and S the separant of A. Then *there exist a non-negative integer* t *and a positive integer* r *such that* $S^t F$ *has a representation*

$$\sum_{j=1}^{r} C_j A^{p_j} A_1^{i_{1j}} A_2^{i_{2j}} \cdots A_{m-l}^{i_{m-l,j}}, \tag{28}$$

with non-negative p_j *and* i_{kj}, *where no two of the* r *sets* $i_{1j}, \cdots, i_{m-l,j}$ *are identical; the* C_j *being forms which are of orders not exceeding* l *in* y_n *and which are not divisible by* A.

For any admissible t, (28) is unique. In what follows, the smallest t will be used.

29. Let F hold the general solution of A. Then, *for the general solution of* A *to be an essential irreducible manifold in the manifold of* F, *it is necessary and sufficient that* (28) *possess a term of the type* $C_jA^{p_j}$, *which term, if* (28) *is considered as a polynomial in* $A, A_1, \cdots, A_{m-l}$, *is of lower degree than every other term of* (28).

The sufficiency proof proceeds as in § 6 of *S. S.*

For the necessity proof, we assume that, among the terms of lowest degree in (28), there is a term which involves derivatives of A. We prove that the general solution of A is not essential.

According to Part I of the present paper, the manifold of F consists of the general solutions of certain forms $B_1, \cdots, B_d$. If there are B_i whose orders in y_n do not exceed l, let T denote the product of such B_i. Otherwise, let $T = 1$. Let T be arranged as a polynomial in the y_{nj} and let U be any coefficient in T. Then U, being a form in $y_1, \cdots, y_{n-1}$, does not hold the general solution of A.

Considering (28) as a polynomial in A and the A_j, we take its terms of lowest degree and select from them those terms which have a highest degree in A_{m-l}. From the terms just taken, we select those for which the degree in A_{m-l-1} is highest. We continue through A_1. Our process isolates a single term of (28), with a definite C_j. This C_j will be used in what follows.

We put $E = UC_jS$. Let $\bar{y}_1, \cdots, \bar{y}_n$ be any solution in the general solution of A which does not annul E. We put, in (28),

$$y_i = \bar{y}_i, \ i = 1, \cdots, n-1; \ y_n = \bar{y}_n + u_0, \tag{29}$$

with u_0 a new unknown. By *S. S.*, § 10, (28) goes over into a form in u_0 which vanishes for $u_0 = 0$ and which has, among its terms of lowest degree in the u_{0i}, terms of order higher than l in u_0.

According to §§ 11-16 of *S. S.*, F has a formal solution

$$y_i = \bar{y}_i, i < n; \ y_n = \bar{y}_n + \phi_1 c + \cdots \tag{30}$$

where ϕ_1 is any solution of a differential equation of order higher than l.

We follow § 17 of *S. S.* Let Σ_i, $i = 1, \cdots, d$, be the closed system whose manifold is the general solution of B_i above. Then every solution (30) is a formal solution of some Σ_i. We say that there is a solution (30) which is a solution of some Σ_i whose B_i has an order higher than l in y_n. Let this be false. Then every solution (30) annuls T above. Under the substitution (29), T goes over into a form $V(u_0)$ which is annulled by every series $\phi_1 c + \cdots$. Because the $\bar{y}_i$ with $i < n$ do not annul U, V is not identically zero. The order of V in u_0 does not exceed l. The proof is now completed as in *S. S.*, § 17.

VI. Equations in Two Unknowns, of the First Order.

GENERALITIES.

30. We deal with an algebraically irreducible form F in the unknowns u and v. F has an order in u and an order in v. We shall assume that the maximum of these two orders is unity.

The manifold of F consists of the general solutions of forms

$$F, B_1, \cdots, B_s. \tag{31}$$

The B_i are determined by the methods of *A. D. E.*, Chapter V, with the help of Part V of the present paper.

Clearing the ground for further operations, we shall show that, if there are B_i in (31), they are of order zero in u and in v. Let us consider B_1. Because F holds the general solution of B_1, the remainder of F with respect to B_1, say for the order u, v of the unknowns, is zero. Thus, the order of B_1 in v cannot exceed unity. If that order were unity, F would be divisible by B_1. Thus, the B_i are simple forms.

31. The problem to which this Part VI is devoted is that of determining the solutions which any B_i, call it B, has in common with the general solution of F.

Fixing our ideas, we assume that B is not free of v. Following § 28, we write

$$S^t F = C_0 B^p + C_1 B^{p_1} B'^{q_1} + \cdots + C_r B^{p_r} B'^{q_r}$$

with S and B' the separant and derivative respectively of B. The orders of the C in v and in u do not exceed 0 and 1 respectively and no C is divisible by B. For every i, $p < p_i + q_i$.

The sufficiency proof of the result in § 29 brings out the fact that every solution which B has in common with the general solution of F is a solution of C_0. Let us suppose that B and C_0 have common solutions. Using the order u, v of the unknowns, let basic sets be obtained for a set of closed irreducible systems whose manifolds make up the manifold of the system C_0, B.

We shall prove that each basic set consists of two forms. The resultant of C_0 and B with respect to v is a non-zero form in u alone, of order at most unity in u. Hence, given any closed irreducible system Σ held by C_0 and B, a basic set of Σ starts with a form U in u, of order at most unity. Because B involves v, B is not divisible by U. Hence the remainder of B with respect to U is not zero. Thus, the basic set of Σ has a second form, of order zero in v, which introduces v.

Let then

$$U, V \tag{32}$$

be a basic set for Σ, one of the closed irreducible systems considered above. U is of order at most unity in u and V is of order zero in v.

It will be proved in § 68 that, *if U is of order unity, the manifold of Σ is contained in the general solution of F.*

As to the case in which U is of order zero, a theorem of E. Gourin [11] shows that if Σ has a solution in common with the general solution of F, the manifold of Σ is contained in that general solution. It thus becomes a question of deciding whether a given solution, $u = \eta$, $v = \zeta$ of Σ is contained in the general solution. As in *S. S.*, § 88, this case can be reduced to the case of $u = v = 0$.

We shall therefore undertake the investigation of the following problem. *Let F vanish for $u = v = 0$. It is required to determine whether $u = v = 0$ is contained in the general solution of F.*

The case of interest, of course, is that in which one or more B_i in (31) vanish for $u = v = 0$.

Through § 67, in which a summary of results is given, we shall be occupied with the problem just stated. Thus, through § 67, F *will vanish for* $u = v = 0$.

ELEMENTS.

32. We consider a relation of the type

$$v = \phi_1 u^{\rho_1} + \cdots + \phi_k u^{\rho_k} + \cdots. \tag{33}$$

The ρ are positive rational numbers with a common denominator, which increase with their subscripts. The ϕ are functions of x, all analytic in some area.[12] It is understood that the second member of (33) may be identically zero. If we differentiate (33) formally, we secure a relation

---

[11] *Bulletin of the American Mathematical Society*, vol. 39 (1933), p. 593.

[12] In the definition of y-solution in *S. S.*, only a common point of analyticity was demanded of the coefficients. The reason for using an area here will appear in § 55.

$$(34) \qquad v_1 = \frac{\partial v}{\partial x} + \frac{\partial v}{\partial u} u_1.$$

The series for $\partial v/\partial u$ may contain a finite number of negative powers of u.

The substitution of the above expressions for v and v_1 into F produces a polynomial in u_1 whose coefficients are series in u which may contain a finite number of negative powers. If this polynomial in u_1 vanishes identically, we shall call the second member of (33) an *element* of F.

For instance, if $F = v_1 + v - u_1 - u$, F has u as an element. Examples of elements can be given, with constant ϕ, which diverge for every $u \neq 0$.

MULTIPLICITIES.

33. Let an element of F be given by (33). If

$$(35) \qquad \frac{\partial F}{\partial v}, \cdots, \frac{\partial^{p-1} F}{\partial v^{p-1}}$$

all vanish identically in x, u, u_1 when v is taken as in (33), while $\partial^p F/\partial v^p$ does not, the element will be said to be of *multiplicity* p.

If F has an element, v or v_1 must figure in F. We shall prove that an element of F has a multiplicity if and only if F, considered as a polynomial in v, is of positive degree.

If F is of zero degree in v, every $\partial^i F/\partial v^i$ vanishes identically. This proves the necessity of the condition. Let the condition be satisfied and let

$$F = \alpha_0 + \alpha_1 v + \cdots + \alpha_n v^n$$

with $n \geqq 1$. Suppose that the n forms $\partial^i F/\partial v^i$, $i = 1, \cdots, n$, all vanish for (33). We have

$$\frac{\partial^n F}{\partial v^n} = n!\, \alpha_n,$$

so that α_n vanishes for (33). Again,

$$\frac{\partial^{n-1} F}{\partial v^{n-1}} = (n-1)!\, \alpha_{n-1} + n!\, \alpha_n v$$

so that α_{n-1} must vanish for (33). Continuing, we find every α_i to vanish for (33). Because F is algebraically irreducible, $\alpha_0, \cdots, \alpha_n$ are relatively prime as polynomials in u, u_1, v_1. Hence, some linear combination of them, with suitable forms for coefficients, is a non-zero form free of v_1, that is, a non-zero form in u. Such a form cannot vanish for (33). This proves the sufficiency.

STRONG SUMS.

34. We consider the effect of making, in F, the substitution

$$v = \phi_1 u^{\rho_1} + \cdot \cdot \cdot + \phi_k u^{\rho_k} \tag{36}$$

where the k numbers ρ are positive and rational, and increase with their subscripts. It is understood that the second member of (36) may be identically zero.

Under this substitution, F goes over into a finite sum

$$\Sigma a_i u^{\alpha_i} u_1^{\beta_i} \tag{37}$$

with each a_i a function of x, each α_i a rational number (possibly negative), and each β_i a non-negative integer.

If (37) is not identically zero, and if, among those of its terms for which $\alpha_i + \beta_i$ is a minimum, there are terms with $\beta_i > 0$, we shall call the second member of (36) a *strong sum* for F.

Example 1. $v^2 - uv + v_1^4 + u_1^3$ has u as a strong sum. In fact, $u + u^\rho$, with ρ any sufficiently large rational number, will be a strong sum.

Example 2. $u + u_1 v_1$ has no strong sum.

From any strong sum of a form F, new strong sums can be derived, as in Example 1, by the addition of terms.

The part played by elements and strong sums will be as follows. Suppose that the manifold of the form u is not contained in the general solution of F. It will turn out that for $u = v = 0$ to be in the general solution, it is necessary and sufficient that F have either a strong sum, or else an element which causes no B_i in (31) to vanish when substituted for v.

INDICES.

35. F is to be as in § 30 and is to admit $u = v = 0$ as a solution. We denote by $\mathfrak{A}$ the region in which the coefficients in F are meromorphic.

Suppose first that F has no strong sum. We shall call a positive integer n the *index* of F if there exists a set of points $\mathcal{E}$, contained in $\mathfrak{A}$ and having no limit point in $\mathfrak{A}$, such that, given any simply connected region $\mathfrak{A}_1$ in $\mathfrak{A}$ which contains no point of $\mathcal{E}$, F has a finite set of distinct elements which satisfy the following conditions:

(a) *The coefficients in the elements are analytic throughout* $\mathfrak{A}_1$.

(b) *The elements have multiplicities and the sum of the multiplicities for the elements of the set is* n.

(c) *Every element of F with coefficients analytic in some area contained in $\mathfrak{A}_1$ coincides with some element of the set.*

It is easy to see that there can be no more than one n as above.

If F has no strong sum and no elements, the index of F will be defined as zero.

If F has a strong sum, the index of F will be defined as ∞.

Our work will show that if F has no strong sum, but has elements, a positive n exists as above. Thus an index will be known to exist for every F which satisfies our assumptions. The index will play a rôle analogous to that of the *y-solution number* in *S. S.*

POLYGONS.

36. We write F in the form

$$\sum_{i=1}^{r} a_i u^{\alpha_i} u_1^{\beta_i} v^{\gamma_i} v_1^{\delta_i} \tag{38}$$

with the a_i functions of x distinct from 0.

We put $\lambda_i = \gamma_i + \delta_i$, $\mu_i = \alpha_i + \beta_i$, and, in a plane referred to rectangular axes, plot the points (λ_i, μ_i). We secure thus r or fewer points, each point associated with one or more terms of F.

We consider those of the plotted points which have a least abscissa—say the abscissa ζ_1—and choose from them that point which has a least ordinate—say the ordinate σ_1. For all points (λ_i, μ_i) with $\lambda_i > \zeta_1$, if such exist, we form the ratio

$$\frac{\sigma_1 - \mu_i}{\zeta_1 - \lambda_i}, \tag{39}$$

which is the slope of the straight segment joining (ζ_1, σ_1) to (λ_i, μ_i). Let us suppose that there are segments whose slopes (39) are negative. Taking those segments whose slope is a minimum, we choose the longest of them. Let its right extremity be denoted by (ζ_2, σ_2).

It may be that there are points with $\lambda_i > \zeta_2$ for which

$$\frac{\sigma_2 - \mu_i}{\zeta_2 - \lambda_i} < 0. \tag{40}$$

If so, we take those points which minimize the first member of (40) and choose from them that point (ζ_3, σ_3) whose abscissa is the greatest.

We continue this construction as long as it is possible to secure segments of negative slope. The polygon formed by the segments obtained will be called the *polygon of F*.

If there are no points with $\lambda_i > \zeta_1$ or if there are no such points for which (39) is negative, the polygon of F is defined as the point (ζ_1, σ_1).

When we speak of the points (λ_i, μ_i) lying on a side of a polygon, the extremities of the side will be included.

Consider a point (λ_i, μ_i), plotted for F, which lies on the polygon of F. If there is a term associated with (λ_i, μ_i) which involves either u_1 or v_1, we shall call (λ_i, μ_i) a *b-point*. If no such term exists, that is, if (λ_i, μ_i) is associated with only a single term of F and that term is free of u_1 and v_1, we shall call (λ_i, μ_i) an *a-point*.

INVESTIGATION OF THE INDEX.

37. We denote the polygon of F by $\mathfrak{P}$. Let (ζ_t, σ_t) be the point of greatest abscissa on $\mathfrak{P}$, that is, the rightmost point on $\mathfrak{P}$. If F is written as in (38), σ_t will be the least of the quantities $\alpha_i + \beta_i$, and ζ_t will be the least value of $\gamma_i + \delta_i$ in those terms of (38) for which $\alpha_i + \beta_i = \sigma_t$.

We are going to work toward the result that F *has an index and the index of* F *is either* ζ_t *or* ∞.

38. We begin by showing that if $\mathfrak{P}$ has a b-point, F has strong sums.

We take first the case in which $\mathfrak{P}$ has at least one side. Let l be some side of $\mathfrak{P}$ on which a b-point lies. Let $-\rho$ be the slope of l. We make, in (38), the substitution

$$v = wu^{\rho} \tag{41}$$

with w an indeterminate which admits of differentiation with respect to x. We have

$$v_1 = w_1u^{\rho} + \rho wu^{\rho-1}u_1. \tag{42}$$

Under (41), a term of (38) associated with a point (λ_i, μ_i) will yield a set of terms, all of degree $\mu_i + \rho\lambda_i$ in u and u_1. If (λ_i, μ_i) is on l, this degree will be the intercept of l on the axis of ordinates. Points not on l produce terms of degree greater than this intercept.

Let A be the sum of those terms of F which correspond to points on l. Under (41), A goes over into a sum

$$\Sigma a_i u^{b_i} u_1^{c_i} \tag{43}$$

with the a_i polynomials in w and w_1. Clearly, if (43) involves u_1 effectively we can fix w as a function ϕ of x so as to make ϕu^{ρ} a strong sum for F.

We shall examine now the case in which (43) is free of u_1. In this case (43) reduces to an expression Bu^q with q the intercept of l and B a polynomial

p14903.

in w and w_1. Now w_1 must appear effectively in B, for (43) goes over into A by the substitution $w = vu^{-\rho}$.

Let C be an irreducible factor of B which involves w_1 and let

$$B = CD. \tag{44}$$

Let ϕ be a function of x which, substituted for w, annuls C but not $\partial C/\partial w_1$. In (44) we put

$$w = \phi + h, \qquad w_1 = \phi_1 + k$$

with ϕ_1 the derivative of ϕ and h and k indeterminates. The members of (44) become identical polynomials in h and k. The terms of lowest degree produced by C are of the first degree and there is a term βk with $\beta \neq 0$. It follows that B produces a polynomial H in which the terms of lowest degree involve k effectively.

Taking any positive rational number δ, we put

$$h = u^{\delta}, \qquad k = \delta u^{\delta-1} u_1$$

Then H goes over into a finite sum of type (37) in which the terms of lowest degree involve u_1.

All in all, if we make in A, above, the substitution

$$v = \phi u^{\rho} + u^{\rho+\delta}, \tag{45}$$

with ϕ as just fixed and δ rational and positive, A goes over into a sum of the type (37) in which the terms of lowest degree involve u_1.

Suppose now that δ is very small. Then the terms yielded by A under (45) will all have degrees close to the intercept of l and will thus have lower degrees than the other terms produced by F under (45). Thus, the second member of (45) is a strong sum for F.

In the case in which $\mathfrak{P}$ consists of a single point which is a b-point, we start with ρ as any positive rational number and the above argument goes through.

39. From this point until the end of § 47, we assume that $\mathfrak{P}$ has no b-point. Two preliminary questions will be treated in this section.

We shall show that 0 is not a strong sum for F. Let this be false. Then F must have terms free of v and v_1 and, in the sum of such terms, u_1 must figure among the terms of lowest degree. This implies that $\mathfrak{P}$ has a b-point, so that our statement is proved.

Suppose that ζ_1 of § 36 exceeds 0. Then $v = 0$ is an element of F. We say that $v=0$ which, according to § 33, has a multiplicity, has the multiplicity ζ_1. It is only necessary to observe, for this, that F has terms free of v_1 and that the minimum of the degrees in v of such terms is ζ_1.

40. Let $\mathfrak{P}$ consist of a single point (ζ_1, σ_1). We say first that F has no strong sum. We know that 0 is not a strong sum. Let the substitution (36) with $\phi_1 \neq 0$ be made in F. The term of F in $u^{\sigma_1}v^{\zeta_1}$ will produce a set of terms of which one, a term in $u^{\sigma_1+\rho_1\zeta_1}$, will have a least degree. The familiar intercept argument shows that this term in u alone will have a lower degree than any other term in the sum (37) which F yields.

What just precedes shows also that F can have no element distinct from 0.

The discussion of § 39 and of the present section proves the theorem of § 37 for the case in which $\mathfrak{P}$ consists of a single point, in particular, for the case of $\zeta_t = 0$.

41. Assuming now that $\mathfrak{P}$ has sides, we undertake to determine the possibilities for ρ_1 and ϕ_1 in a strong sum, or in an element distinct from 0.

We shall prove that ρ_1 is the negative of the slope of some side of $\mathfrak{P}$. Let this be false. Let the vertices of $\mathfrak{P}$, arranged according to increasing abscissas, be

$$(\zeta_1, \sigma_1), \cdots, (\zeta_t, \sigma_t). \tag{46}$$

If there are sides of $\mathfrak{P}$ of slopes greater than $-\rho_1$, let (ζ_j, σ_j) be the first point from the left in (46) which is the extremity of such a side. Otherwise let $j=t$. We consider a line through (ζ_j, σ_j) of slope $-\rho_1$. Then all points plotted for F other than (ζ_j, σ_j) lie above this line. It follows, as in § 40, that a substitution $v = \phi_1 u^{\rho_1} + \cdots$ in F produces a non-zero term, free of u_1, which is of lower degree than any other term produced. This proves our statement.

42. Let l be a side of $\mathfrak{P}$ and let $-\rho_1$ be the slope of l. We make in F the substitutions (41), (42), with $\rho = \rho_1$. The terms of F associated with points on l produce, collectively, an expression $L_1(w)u^\tau$ where L_1 is a polynomial in w, and τ the intercept of l on the axis of ordinates. The degree of L_1 is the abscissa of the right extremity of l. The remaining terms of F produce, collectively, an expression M_1 which, arranged as a sum of power products of u and u_1, will have all its terms of degree greater than τ in u and u_1.

Clearly, if $\phi_1 u^{\rho_1}$ is to be the first term of an element or of a strong sum, we must have

$$L_1(\phi_1) = 0. \tag{47}$$

43. Let ϕ_1 be any solution of (47) distinct from zero.[13] We make, in F, the substitution

$$v = \phi_1 u^{\rho_1} + v'. \tag{48}$$

Then F goes over into an expression F' in v' and u. F' will be a polynomial in v' and v'_1, with coefficients which are sums of power products of u and u_1. The exponents of u_1 in the coefficients will be non-negative integers; those of u will be rational numbers and some of them may be negative.

We write F' in the form (38). Now, however, the α_i may be fractional, and even negative. We form a polygon for F' in the manner explained in § 36. We denote this polygon by $\mathfrak{P}'$. We are going to study $\mathfrak{P}'$.

A term in F in

$$u^\alpha u_1^\beta v^\gamma v_1^\delta \tag{49}$$

associated with the point $(\gamma + \delta, \alpha + \beta)$, contributes to F' terms coming from

$$u^\alpha u_1^\beta (v' + \phi_1 u^{\rho_1})^\gamma (v'_1 + \phi'_1 u^{\rho_1} + \rho_1 \phi_1 u^{\rho_1 - 1} u_1)^\delta \tag{50}$$

with ϕ'_1 the derivative of ϕ_1. Let us consider any term coming from (50). If its degree in v' and v'_1 is $\gamma + \delta - a$ with $0 \leqq a \leqq \gamma + \delta$, its degree in u and u_1 will be $\alpha + \beta + \rho_1 a$. Such a term will be associated with the point

$$(\gamma + \delta - a, \ \alpha + \beta + \rho_1 a). \tag{51}$$

One of the points (51) will be $(\gamma + \delta, \alpha + \beta)$. The others lie on a line sloping upward from that point, with slope $-\rho_1$.

Now, let h represent the point which is the right extremity of l. What precedes makes it geometrically obvious that h and all plotted points on $\mathfrak{P}$ to the right of h are points plotted for F' and, indeed, the lowest plotted points of their respective abscissas. Thus, h is a vertex of $\mathfrak{P}'$ and $\mathfrak{P}$ and $\mathfrak{P}'$ coincide from h onward to the right, that is, they have the same plotted points, which are a-points for both of them.

We shall now examine $\mathfrak{P}'$ to the left of h. Let ϕ_1 be a solution of (47) of multiplicity p. We shall call p the *multiplicity* of ϕ_1. Under the substitution (48), the terms of F associated with points on l will produce collectively the expression[14]

$$L_1(\phi_1 + u^{-\rho_1} v') u^\tau. \tag{52}$$

The term of highest degree in v' in (52) will be the term of F' associated with h. The other terms of F' coming from (52) will yield points which lie on l or on l produced to the left. The lowest power of v' which figures effectively in (52) is the p-th power. This means that F' has a term in $u^{\tau-\rho_1 p}v'^p$ which is the only term in F' associated with the point $(p, \tau-\rho_1 p)$ and that every other point, plotted for F', of abscissa p, has an ordinate greater than $\tau-\rho_1 p$. Furthermore, all points of abscissa less than p which may be plotted for F' lie in the interior of the upper half-plane determined by l.

---

[13] The number of such solutions is the length of the horizontal projection of l.

[14] Put $w = \phi_1 + u^{-\rho_1} v'$.

It follows that if p is less than the abscissa of h, $\mathfrak{P}'$ has a side of slope $-\rho_1$ which joins $(p, \tau-\rho_1 p)$ to h, and that this side has only a-points.

Whether or not p is less than the abscissa of h, if there are points plotted for F' of abscissa less than p, $\mathfrak{P}'$ has sides of slope less than $-\rho_1$ and $(p, \tau-\rho_1 p)$ is the rightmost extremity of the rightmost such side. Any b-points which $\mathfrak{P}'$ may have lie on sides of slope less than $-\rho_1$.

44. If $\mathfrak{P}'$ has a b-point, lying on a side of slope $-\rho_2 < -\rho_1$, an application of the method of § 38 shows that F has a strong sum of one of the two forms

$$\phi_1 u^{\rho_1} + \phi_2 u^{\rho_2}, \qquad \phi_1 u^{\rho_1} + \phi_2 u^{\rho_2} + u^{\rho_2+\delta}.$$

45. We assume now that $\mathfrak{P}'$ has no b-points. Let q be the abscissa of the leftmost vertex of $\mathfrak{P}'$.

For F' to be annulled by $v'=0$, that is, for $\phi_1 u^{\rho_1}$ to be an element of F, it is necessary and sufficient that q exceed 0.

Suppose that $q>0$. Then $\partial^j F'/\partial v'^j$ vanishes for $j<q$ but not for $j=q$. This shows that $\phi_1 u^{\rho_1}$ is an element of F of multiplicity q.

If $q=0$, we show as in § 39 that $\phi_1 u^{\rho_1}$ is not a strong sum for F.

46. Suppose now that F has an element with at least two terms or a strong sum with at least two terms, the first term, in either case, being $\phi_1 u^{\rho_1}$ above. The arguments of §§ 41, 42 show that $-\rho_2$ is the negative of the slope of a side of $\mathfrak{P}'$ and that ϕ_2 is a root of a certain equation

$$L_2(\phi)=0 \tag{53}$$

in which the number of solutions distinct from zero, which is the length of the horizontal projection of a side of $\mathfrak{P}'$, does not exceed p of § 43.

47. Let us assume, then, that $\mathfrak{P}'$ has sides of slope less than $-\rho_1$. Let $-\rho_2$ be the slope of some such side. We form the corresponding equation (53) and choose any solution ϕ_2 of (53) which is distinct from 0. We put in F'

$$v'=\phi_2 u^{\rho_2}+v'',$$

whereupon F' goes over into an expression F'' in v''. We form, as above, a polygon $\mathfrak{P}''$ for F''. If p_1 is the multiplicity[15] of ϕ_2, $\mathfrak{P}''$ will have an a-point—call it h—of abscissa p_1 and, if $\mathfrak{P}''$ has sides of slope less than $-\rho_2$, h will be the rightmost extremity of the rightmost such side. If there are no such sides, h is the leftmost point on $\mathfrak{P}''$. Such b-points as $\mathfrak{P}''$ may have will lie on sides of slope less than $-\rho_2$.

48. We conclude the proof of the result stated in § 37.

If $\mathfrak{P}$ has a b-point, the index of F is ∞.

Suppose that $\mathfrak{P}$ has no b-point. Let r_1 stand for ζ_1 in § 41. By § 39, F will have r_1 (possibly 0) zero elements. There will perhaps be certain possibilities for terms $\phi_1 u^{\rho_1}$ of other elements or of strong sums. The sum of r_1 and of the multiplicities of the ϕ_1 is ζ_t of § 37.

For each $\phi_1 u^{\rho_1}$ we find an F'. If some F' yields a b-point, the index of F is ∞. If no b-points are met, we proceed with each F' as in §§ 45, 47. We find that F has a certain number $r_2 \geqq r$ of zero elements and elements $\phi_1 u^{\rho_1}$ and also, perhaps, a certain number of possibilities $\phi_1 u^{\rho_1} + \phi_2 u^{\rho_2}$ for the beginnings of strong sums or of other elements. The sum of r_2 and of the multiplicities of the ϕ_2 is ζ_t.

At the third step, we form an F'' for each $\phi_2 u^{\rho_2}$. We continue in this manner. There are two ways in which our process, having been carried through k steps, may terminate at the $(k+1)$-th step. Firstly, we may meet an $F^{(k)}$ with a b-point. In that case, the index of F is ∞. Secondly, it may be that no $F^{(k)}$ has a polygon with a side of slope less than the negative of the ρ_k associated with that $F^{(k)}$. In that case, F will have no strong sum and will have precisely ζ_t elements of the types 0 or

$$\phi_1 u^{\rho_1} + \cdots + \phi_h u^{\rho_h}, \qquad h \leqq k,$$

in harmony with § 37.

Let us assume that the process does not terminate in a finite number of steps. Then F has no strong sum and, from some step on, no new finite elements appear. That is, if k is large, we will, in the first k steps, have isolated a fixed number r of finite elements and there may be in addition a finite number of possibilities.

$$\phi_1 u^{\rho_1} + \cdots + \phi_k u^{\rho_k} \tag{54}$$

for the beginnings of elements with an infinite number of terms. The sum of r and of the multiplicities of the ϕ_k, for every large k, is ζ_t. After the finite elements have been isolated, the number of distinct expressions (54) cannot decrease as k increases. Thus, after a certain step, there will be a fixed number of distinct expressions (54) and, when an $F^{(k)}$ is determined for any of these expressions, we get a ϕ_{k+1} with the same multiplicity as ϕ_k. An $F^{(k)}$ with k large does not vanish for $v^{(k)} = 0$ and its polygon has just one side of slope less than $-\rho_k$.

---

[15] Defined as for ϕ_1.

This means that we are forming a certain number of infinite series which may be elements of F. Let

$$(55) \qquad \phi_1 u^{\rho_1} + \cdots + \phi_k u^{\rho_k} + \cdots$$

be any one of these infinite series. Let p be the common multiplicity of the ϕ_k, in (55) with k large. We shall prove that (55) is an element of F of multiplicity p.

First, we shall show that the ρ_k in (55) have a common denominator. Let, for some large k,

$$(56) \qquad L_k(\phi) = 0$$

be the equation, similar to (53), which determines ϕ_k. Then the degree of L_k in ϕ is p and ϕ_k is a root of (56) of multiplicity p. Thus

$$(57) \qquad L_k(\phi) = a(\phi - \phi_k)^p$$

with a a function of x. This means that the side of $\mathfrak{P}^{(k-1)}$ of slope $-\rho_k$ has on it points, plotted for $F^{(k-1)}$, of each of the abscissas $0, 1, \cdots, p$. If h_0 is that one of these points whose abscissa is 0 and if h_1 is the point of abscissa 1, ρ_k will be the difference of the ordinates of h_0 and h_1. These ordinates are linear combinations of $\rho_1, \cdots, \rho_{k-1}$ and unity, with integral coefficients. Hence ρ_k is such a linear combination. Thus we can use, for the denominator of ρ_k, the common denominator of $\rho_1, \cdots, \rho_{k-1}$.

We prove now that (55) is an element of F of multiplicity p. Let s be the ordinate of the point on $\mathfrak{P}^{(k)}$ of abscissa p. As seen above, s is independent of k for k large.

For any large k, let polygons be formed in the usual manner for the p expressions $\partial^j F^{(k)}/\partial v^{(k)j}$, $j = 0, \cdots, p$. The discussion of (57) shows that the leftmost point of such a polygon will be on the axis of ordinates and will have for ordinate

$$(58) \qquad s + (p - j)\rho_{k+1}.$$

For $j = p$, (58) equals s for every large k, but, for $j < p$, (58) becomes infinite with k. For any j, (58) is the lowest of the degrees in u, u_1 of the terms in the expression obtained by replacing $v^{(k)}$ in $\partial^j F^{(k)}/\partial v^{(k)j}$ by 0. The same expression is obtained on replacing v in $\partial^j F/\partial v^j$ by the sum of the first

k terms in the second member of (55). This shows that (55) is an element of F of multiplicity p.

We have thus established the result stated in § 37.[16]

MULTIPLICITIES AND VANISHING DERIVATIVES.

49. Suppose that (31) contains certain B_i, say

$$B_1, \cdots, B_d \tag{59}$$

which involve v effectively and which vanish for $u = v = 0$. Let B be any of the B_i in (59).

Referring to Part V, we write

$$S^a F = C_0 B^p + C_1 B^{p_1} B'^{q_1} + \cdots + C_r B^{p_r} B'^{q_r}, \tag{60}$$

with S and B' respectively the separant and derivative of B. The C are forms of order zero in v. Furthermore,

$$p > 0; \; p_i + q_i > p, \qquad (i = 1, \cdots, r).$$

Let a given relation (33) imply $B = 0$. Then the second member of (33) is an element of F. One proves as in *S. S.*, § 55, that (33) gives an element of F of multiplicity p. Furthermore, (33) implies

$$\frac{\partial^{l_1+l_2} F}{\partial v^{l_1} \partial v_1^{l_2}} = 0$$

for $l_2 > 0$, $l_1 + l_2 \leqq p$.

FINAL CRITERIA.

50. We develop now a test for determining whether F has either a strong sum, or an element which annuls no B_i in (59) when substituted for v.

Let each B_i in (59) be written as a polynomial in v and let v^{q_i} be the lowest power of v in B_i whose coefficient is not divisible by u. The equation $B_i = 0$ has q_i solutions of the type (33). Let p_i be the value of p in (60) for $B = B_i$. Let

$$m = p_1 q_1 + \cdots + p_d q_d. \tag{61}$$

We compare m with ζ_t.

Suppose that $\zeta_t > m$. Then F has either a strong sum, or an element which annuls no B_i in (59).

Suppose that $\zeta_t < m$. Then ζ_t cannot be the index of F, so that F has a strong sum.

---

[16] The matter of the areas in § 35 is handled as in *S. S.*, § 54.

Suppose that $\zeta_t = m$. We shall show, in what follows, how to determine whether the index of F is ζ_t or ∞. If the index is ζ_t, F has no strong sum and every element annuls some B_i.

51. We consider the $q_1 + \cdots + q_d$ solutions of type (33) of the relations $B_i = 0$, $i = 1, \cdots, d$. Let k be a positive integer such that no two of these solutions coincide through their first k terms.

We examine the process of §§ 38-48 for finding elements and strong sums of F.

Let g be any positive integer not greater than k. If the process terminates at the g-th step, it must be either that we have encountered a $\mathfrak{P}^{(g-1)}$ with b-points or that no $\mathfrak{P}^{(g-1)}$ has a side of slope less than the corresponding $-\rho_{g-1}$. The manner of termination would indicate whether the index is ∞ or ζ_t.

Let us suppose that the process does not terminate at the k-th step or at an earlier step. We shall prove that the index is ζ_t.

The non-termination means that we have met certain $\mathfrak{P}^{(k-1)}$ which have no b-points and have sides of slopes less than the associated $-\rho_{k-1}$. There will have been isolated a certain number r_k of elements which are either zero or possess at most $k-1$ terms. The sum of r_k and of the multiplicities of the ϕ_k which the $F^{(k-1)}$ yield is ζ_t.

Thus, $r_k < \zeta_t = m$. Then the B_i must have solutions (33) with at least k terms and the ρ_k of the k-th terms must be negatives of slopes of sides of the $\mathfrak{P}^{(k-1)}$.

Let

$$\phi_1 u^{\rho_1} + \cdots + \phi_k u^{\rho_k} + \cdots, \tag{62}$$

consisting of at least k terms, annul some B_i, say B_1. Let (62) have multiplicity p for F. Then ρ_k in (62) is the negative of the slope of a side of the polygon for some $F^{(k-1)}$ which, for the substitution

$$v^{(k-1)} = \phi_k u^{\rho_k} + v^{(k)},$$

yields an $F^{(k)}$ which is annulled by

$$\phi_{k+1} u^{\rho_{k+1}} + \cdots. \tag{63}$$

One proves now, following *S. S.*, §§ 61, 62, that, if the polygon of $F^{(k)}$ has sides of slope less than $-\rho_k$, p is the abscissa of the right extremity of the rightmost such side. If there are no such sides, p is the least abscissa of the points plotted for $F^{(k)}$. One sees also, as in *S. S.*, § 63, that (63) is zero if and only if p is the least abscissa of the points plotted for $F^{(k)}$.

We suppose that (63) is not zero, so that the polygon $\mathfrak{P}^{(k)}$ of $F^{(k)}$ has at least one side of slope less than $-\rho_k$. We shall prove that $\mathfrak{P}^{(k)}$ has a single such side, namely, a side which has slope $-\rho_{k+1}$ (as in (63)) and has its left end on the axis of ordinates. It will be seen also that this side has no b-points.

Let l be the rightmost side of $\mathfrak{P}^{(k)}$ whose slope is less than $-\rho_k$. As in *S. S.*, § 64, it is seen that the slope of l is not less than $-\rho_{k+1}$.

Suppose now that l has a b-point and let the rightmost such b-point be denoted by h_1. The abscissa of h_1 is less than p.

Suppose first that one of the terms associated with h_1 involves $v_1^{(k)}$; let it be a term in

(64) $$v^{(k)\,l_1} v_1^{(k)\,l_2}$$

with $l_2 > 0$. Let

$$K = \frac{\partial^{l_1+l_2} F^{(k)}}{\partial v^{(k)\,l_1} \partial v_1^{(k)\,l_2}}.$$

By § 49, K is annulled by (63). We form a polygon for K. A point for $F^{(k)}$ yields, for K, a point $l_1 + l_2$ units to the left, or no point, according as the point for $F^{(k)}$ does or does not have a term associated with it which is divisible by (64). As in *S. S.*, § 64, we see that (63) does not annul K.

Suppose now that the terms of $F^{(k)}$ associated with h_1 are all free of $v_1^{(k)}$. Then u_1 appears in those terms. Denoting by q the abscissa of h_1, we let

$$K = \frac{\partial^q F^{(k)}}{\partial v^q}.$$

Then a point plotted for $F^{(k)}$ yields, for K, a point q units to the left, or no point, according as the point for F is or is not associated with a term divisible by v^q. In particular, h_1 yields a point h_2 on the axis of ordinates and the right end of l yields a point which, when joined to h_2, produces a leftmost side, call it l_1, for the polygon of K. The slope of l_1 is that of l and hence is not less than $-\rho_{k+1}$. When (63) is substituted into K, there will result, from the point h_2, terms involving u_1 whose degree in u and u_1 is the ordinate of h_2. The only other terms of K which can conceivably produce terms of this degree are those associated with the points on l_1 other than h_2. Such points are a-points and cannot yield terms which involve u_1. Thus, (63) does not annul K, so that l has no b-point.

The discussion is continued as in *S. S.*, § 64.

An argument like that in *S. S.*, § 65, shows that the index of F is ζ_t. Our discussion has taken care of a set of $F^{(k)}$ for which the sum of the multiplicities of the ϕ_k is at least $\zeta_t - r_k$. This shows that all $F^{(k)}$ are accounted for.

SUFFICIENT CONDITIONS.

52. We denote the general solution of F by $\mathfrak{M}$.

Let F have an element, given by (33). We calculate the successive derivatives of v in (33) formally, expressing each derivative as a polynomial in the derivatives of u with series in u as coefficients.

Suppose that every form which holds $\mathfrak{M}$ vanishes identically in $x, u, u_1, \cdots$, when $v, v_1, \cdots$ are replaced in the form by their expressions found from (33). We shall say, in that case, that the element given by (33) *is in* $\mathfrak{M}$ or *belongs to* $\mathfrak{M}$.

We say that if F has an element which belongs to $\mathfrak{M}$, the solution $u = v = 0$ is in $\mathfrak{M}$. In short, if a form G possesses a term free of the u_i, v_i, the substitution of v as in (33) into G cannot produce zero.

Let us prove that if there are B_i as in (59), no element in $\mathfrak{M}$ can annul any B_i. Let B be any such B_i. Referring to § 31, we denote by R the resultant of B and C_0 with respect to v. Then R, which is a non-zero form in u, vanishes for all solutions of B which are contained in $\mathfrak{M}$. If $G_1, \cdots, G_q$ is a finite set of forms whose manifold is $\mathfrak{M}$, some power of R is a linear combination of $B, G_1, \cdots, G_q$ and their derivatives. Any element in $\mathfrak{M}$ which annuls B would thus annul R. As R does not involve v, R cannot vanish for such an element.

We prove now that an element of F which annuls no B_i in (59) is in $\mathfrak{M}$. Referring to (31), we let

$$T = B_1 B_2 \cdots B_s.$$

If Q is any form which holds $\mathfrak{M}$, QT holds F. It follows that every element of F which does not annul T annuls Q. Now a B_i in (31) which is annulled by an element must involve v and must vanish for $u = v = 0$. Thus an element which annuls no B_i in (59) does not annul T and hence is in $\mathfrak{M}$.

Thus, for $u = v = 0$ to be in $\mathfrak{M}$, it is sufficient for F to have an element which belongs to $\mathfrak{M}$. If F has elements and if there are no B_i as in (59), every element of F is in $\mathfrak{M}$. If there are B_i in (59), an element of F is in $\mathfrak{M}$ if and only if it annuls no B_i.

53. We shall prove now that if F has a strong sum, $u = v = 0$ is in $\mathfrak{M}$. Let F have a strong sum given by (36). We arrange the sum (37) as a polynomial in u_1 and then equate it to zero. We secure an equation

$$(65) \qquad A_0 + A_1 u_1 + \cdots + A_q u_1^q = 0$$

where the A are sums of terms of the form bu^ρ, with b a function of x, and ρ

rational. Regarding (65) as an algebraic equation for u_1, we use the Newton polygon method to form for it solutions of the type

$$(66) \qquad u_1 = \gamma_1 u^{\sigma_1} + \gamma_2 u^{\sigma_2} + \cdots$$

with the γ and σ as usual. Because u_1 figures in the terms of lowest degree in (37), there will be at least one solution (66), (possibly identically zero), with $\sigma_1 \geqq 1$. We consider such a solution (66) and denote by p the common denominator of its σ. Considering (66) as a differential equation, we put $u = w^p$. Then (66) goes over into a differential equation

$$(67) \qquad w_1 = f(x, w)$$

with f analytic for x in some area and w small. Furthermore, $f(x, 0)$ is zero for every x. Equation (67) admits a one-parameter family of solutions

$$w = \psi(x, c)$$

where ψ is analytic for x in some area and for c small, and where ψ, without vanishing identically in x and c, vanishes identically in x for $c = 0$. Representing the p-th power of ψ by ζ, we see that (37) vanishes for $u = \zeta$ if c is small and distinct from 0. For x in a suitable area, we obtain, on replacing u by ζ in (36), a one-parameter family of analytic functions v which tend uniformly towards zero as c approaches zero. We have thus a one-parameter family of solutions of F which approach uniformly, as c decreases, the solution $u = v = 0$. For c small these solutions cannot annul any B in (31). This is because the second member of (36) is not an element of F. Hence, for c small, we get solutions in $\mathfrak{M}$. This proves that $u = v = 0$ is in $\mathfrak{M}$.

STATEMENT OF NECESSARY CONDITIONS.

54. In the sections which follow, we shall prove that if $u = v = 0$ is in $\mathfrak{M}$, at least one of the following conditions is satisfied:

(a) *The manifold of the form u belongs to $\mathfrak{M}$.*
(b) *F has an element which belongs to $\mathfrak{M}$.*
(c) *F has a strong sum.*

That the satisfaction of any one of these conditions insures the presence of $u = v = 0$ in $\mathfrak{M}$ is already known to us.

A Normalization.

55. It may be necessary later to interchange the letters u and v in F. We represent F by $F(u, v)$. When u and v are interchanged we secure a form which we shall denote by $F(v, u)$.

We are going to prove that if one of the three conditions of § 54 is satisfied by $F(u, v)$, then some one of those conditions is satisfied by $F(v, u)$.

If (a) is satisfied, $F(v, u)$ has the element 0, which belongs to its general solution.

Let (b) be satisfied. If the element described is zero, the general solution of $F(v, u)$ contains the manifold of $u = 0$. If not, the inversion of (33) produces an element of $F(v, u)$ in the general solution of $F(v, u)$. The fact that the coefficients in (33) have a common area of analyticity permits the inversion to be made.

We prove now that if $F(u, v)$ has a strong sum, $F(v, u)$ has a strong sum. Let $F(u, v)$ have a strong sum given by (36). We know from earlier work that if the second member of (36) is zero, then $v = u^\rho$ with ρ sufficiently large and rational is a strong sum for $F(u, v)$. We assume in what follows that the second member of (36) is not zero.

For v as in (36) and for u_1 as in (66), $F(u, v)$ vanishes identically in u and x. Equation (36) defines u as a power series in v,

$$u = \psi_1 v^{\gamma_1} + \psi_2 v^{\gamma_2} + \cdots \tag{68}$$

with $\gamma_1 > 0$. We replace u in (66) by the series in (68). We replace u_1 in (66) by its expression found from (68) by differentiation. This gives a relation between v_1 and v which reduces to a form

$$v_1 = \zeta_1 v^{\delta_1} + \zeta_2 v^{\delta_2} + \cdots \tag{69}$$

with $\delta_1 \geqq 1$. Then $F(u, v)$ vanishes identically in v and x for u as in (68) and v_1 as in (69). On the other hand, because the second member of (36) is not an element, $F(u, v)$ does not vanish identically in x, v, v_1 for u as in (68). When the substitution (68) is made in $F(u, v)$, $F(u, v)$ goes over into a polynomial

$$A_0 + A_1 v_1 + \cdots + A_q v_1{}^q \tag{70}$$

with the A series in v. Because (70) vanishes when v_1 is taken as in (69), the Newton polygon for (70) must have a side of slope not greater than -1. If, instead of substituting into $F(u, v)$ the entire series in (68), we substitute a sufficiently long segment of the series, we secure an expression similar to

(70) with the same Newton polygon which (70) has; this is because the new expression differs from (70) by terms of high degree in v.

This means that if we take a sufficiently long segment of the series in (68) and replace v in the segment by u, we obtain a strong sum for $F(v, u)$.

56. Continuing with F described as in § 30, we make the assumption that $u = v = 0$ belongs to $\mathfrak{M}$. Suppose that there exists a form, holding $\mathfrak{M}$, of the type

$$u^p + A$$

where A either is identically zero or else has each of its terms of degree higher than p in the u_i, v_i. By Part III, there cannot exist a form holding $\mathfrak{M}$ which is of the type

$$v^q + B$$

where B either is identically zero or else has each of its terms of degree higher than q in the u_i, v_i.

57. Let F be as in § 30 with $u = v = 0$ in $\mathfrak{M}$.

Suppose that F does not involve u. Then $v = 0$ is in the general solution of F considered as a form in v alone. Thus, zero is an element of F in $\mathfrak{M}$ and (b) of § 54 is satisfied. If F does not involve v, (a) is satisfied.

58. On the basis of §§ 55-57, we assume that F, described at the start as in § 30, has $u = v = 0$ in its general solution $\mathfrak{M}$, that F involves both u and v effectively[17] and that no form $u^p + A$ as in § 56 holds $\mathfrak{M}$. It will be proved, in what follows, that one of (b) and (c) of § 54 is satisfied.

NECESSITY PROOF.

59. We denote by S the separant of F for the order v, u of the unknowns.[18] Let p be any positive integer. We consider the forms

$$F, F_1, \cdots, F_p \tag{71}$$

where F_i is the i-th derivative of F.

We now regard the F in (71) as simple forms. The unknowns u_i in the F will be $u_0, \cdots, u_{p'}$, where p' is p or $p + 1$ according as the order of F in u is 0 or 1. The unknowns v_i are $v_0, \cdots, v_{p''}$ with p'' either p or $p + 1$.

---

[17] The assumption that F involves v is made principally for definiteness of procedure in § 59.

[18] Note that F involves u.

The totality of simple forms which vanish for all solutions of (71) with $S \neq 0$ is a prime system Λ. Λ has the solution $u_i = 0$, $i = 0, \cdots, p'$; $v_i = 0$, $i = 0, \cdots, p''$.

In Λ, we replace each u_i with $i > 0$ by $u_0 z_i$, with z_i a new unknown. We replace v_i, $i = 0, \cdots, p''$ by $u_0 w_i$. Then Λ goes over into a system Ξ in u_0, the z and w. As Λ has solutions with $u_0 \neq 0$, Ξ also has such solutions. The totality Ω of forms in u_0, the z and w which vanish for all solutions of Ξ with $u_0 \neq 0$ is a prime system.

We shall prove that Ω has solutions with $u_0 = 0$. Let this be false. Let $G_1, \cdots, G_q$ be a finite subset of forms of Ω with the same manifold as Ω. Then

$$u_0, G_1, \cdots, G_q$$

has no solutions, so that there exists a relation

$$1 = K_0 u_0 + K_1 G_1 + \cdots + K_q G_q.$$

Then $1 - K_0 u_0$ is a form in Ω. We prove as in *S. S.*, § 69, that Λ has a form M of the type $u_0^g + A$ with each term of A of degree higher than g in the u_i, v_i. M as a differential polynomial holds $\mathfrak{M}$. This contradicts § 58. Thus Ω has solutions with $u_0 = 0$.

60. According to *S. S.*, § 68, there exists a set of functions $\phi(x, h)$, one for each unknown in Ω, analytic for x in some area and for h small, the ϕ for u_0 reducing to 0 for $h = 0$, which annul the forms in Ω but not $u_0 S$.

Passing from Ω to Λ, we have an analogous set of $\phi(x, h)$ for the unknowns in Λ. Let r be the lowest power of h in the expansion of the ϕ which corresponds to u_0. Then the expansion of any other ϕ which is not zero has a least power of h no less than r.

61. Let Φ be some finite system of differential polynomials in u and v, containing F, whose manifold is $\mathfrak{M}$.

We select a value a of x such that:

I. The coefficients in Φ are analytic at a.

II. Every $\phi(x, h)$ for Λ, as in § 60, is analytic for $x = a$ and for h small.

III. The coefficient of h^r in the ϕ corresponding to u_0 does not vanish at a.

IV. S does not vanish identically in h when its unknowns are replaced by their ϕ and x is put equal to a.

62. When x is replaced by a in the ϕ, the ϕ go over into functions of h

which are analytic for h small. We denote the functions of h associated with the u_i by the generic symbol α and those associated with the v_i by β. The α for u_0 has a zero of order r for $h = 0$. Every other α which is not identically zero, and every β which is not identically zero, has a zero of order at least r for $h = 0$. For $x = a$, the F in (71) vanish identically in h when the unknowns are replaced by their α, β.

Let F_i denote for $i > p$, as above for $i \leqq p$, the i-th derivative of F. In F_{p+1}, we replace x by a, the unknowns other than $u_{p'+1}$ and $v_{p''+1}$ by their α and β, and $v_{p''+1}$ by 0. Equating F_{p+1} to zero after these substitutions, we secure a linear equation for $u_{p'+1}$. Because of IV of § 61, $u_{p'+1}$ is determined as a function of h which is either analytic, or else has a pole, for $h = 0$. We treat F_{p+2} similarly, making the substitutions described above and, furthermore, replacing $v_{p''+2}$ by 0 and $u_{p'+1}$ by the function of h found above. We find, for $u_{p'+2}$, a function of h which is either analytic, or has a pole, for $h = 0$.

The process just described determines a one-parameter family of solutions in $\mathfrak{M}$,

$$v = \beta_0(h) + \beta_1(h)(x-a) + \cdots + \frac{\beta_{p''}(h)}{p''!}(x-a)^{p''}, \tag{72}$$

$$u = \alpha_0(h) + \cdots + \frac{\alpha_k(h)}{k!}(x-a)^k + \cdots. \tag{73}$$

The β, and the α_k with $k \leqq p'$, have zeros of orders at least r for $h = 0$. The α_k with $k > p'$ may conceivably have poles for $h = 0$.

63. Let

$$w = \alpha_0(h) + \cdots + \frac{\alpha_{p'}(h)}{p'!}(x-a)^{p'}. \tag{74}$$

Let q be the highest of the orders of the derivatives of u and v which appear in Φ of § 61. We assume that p of § 59 is taken greater than q.

When the second members of (72) and (74) are substituted for v and u respectively in any form G of Φ, G goes over into a series

$$\gamma = \sum_{k=r}^{\infty} \psi_k(x) h^k \tag{75}$$

with the ψ analytic for $x = a$. We prove as in *S. S.*, § 71, that every ψ which is not identically zero has a zero at a of order at least $p - q$.

Representing w in (74) by u, we find, as in *S. S.*, § 72, that u satisfies, for h small, a differential equation

$$u_1 = \mu_0 u + \mu_1 u^{1+1/r} + \cdots + \mu_k u^{1+k/r} + \cdots \tag{76}$$

with μ which are functions of x, analytic for $x = a$. With this same meaning for u, we find for v as given by (72) an expansion

$$v = \nu_0 u + \nu_1 u^{1+1/r} + \cdots + \nu_k u^{1+k/r} + \cdots \tag{77}$$

with ν which are analytic for $x = a$.

64. Following *S. S.*, § 73, we find from (76) and (77), by differentiation, expansions in powers of $u^{1/r}$ for u_i with $i \leqq q$ and v_i with $i \leqq q$ where q is as in § 63. These expansions contain no power of u lower than the first power.

Let these expansions be substituted into any form G in Φ. Then G becomes an expression

$$\zeta_0 u + \zeta_1 u^{1+1/r} + \cdots . \tag{78}$$

As in *S. S.*, § 73, we prove that every ζ which is not identically zero has a zero at a of order at least $p - q$.

As in *S. S.*, § 74, it is possible to find a value a of x which can be used for a sequence of values of p increasing to ∞.

65. Referring to the necessity conditions of § 54, we assume that F, described as in § 58, has no strong sum. We find in §§ 65, 66 an element which belongs to $\mathfrak{M}$.

We take over the discussion of expansions in §§ 76-81 of *S. S.*, speaking here of u-expansions.

Choosing a sequence of values of p which increase towards ∞, we form, for each p, a pair of series of the types shown in (76) and (77). Without loss of generality, we assume that these sequences of u and v, and also the sequence of v_1 which they give by differentiation, have strong characteristics.

When v and u_1 are replaced in F by the corresponding series in (76) and and (77), we secure from F, according to § 64, a sequence of u-expansions converging to zero. We are going to extract from the sequence of v a subsequence which converges to an element of F. Later it will be shown that this element is in $\mathfrak{M}$.

Suppose first that the v converge to zero. We shall prove that zero is an element of F. Let this be false. Then F has terms free of v and v_1. Because F has no strong sum, $\mathfrak{P}$ has no b-points. Thus, among the terms free of v and v_1, there is a term of the type αu^q which is of lower degree than any of the other terms. When v and u_1 are replaced in F by their u-expansions, every term of F which involves v or v_1 produces a sequence of u-expansions converging to zero. The terms free of v and v_1, other than αu^q, produce sequences of characteristics exceeding q. This shows that zero is an element of F.

Now, suppose that the sequence of v has a finite characteristic ρ_1. Of course, $\rho_1 \geqq 1$. We prove, as in *S. S.*, § 84, that $\mathfrak{P}$ has sides and that $-\rho_1$ is the slope of a side. Continuing as in *S. S.*, we find a subsequence of values of p, and a first term $\phi_1 u^{\rho_1}$ of an element of F, such that, for the p of the subsequence, the expansions $v - \phi_1 u^{\rho_1}$ form a sequence of characteristic exceeding ρ_1.

Let F go over into an expression F' in v' and u under the substitution (48). Then the substitution of v and u_1, as in (77) and (76), into F produces the same u-expansions as the substitution of $v' = v - \phi_1 u^{\rho_1}$ and u_1 into F'. Using the fact that the polygon of F' has no b-points, we prove that, if the v' converge to zero, $\phi_1 u^{\rho_1}$ is an element of F. If the sequence of v' has a finite characteristic $\rho_2 > \rho_1$, we continue as in *S. S.*, § 85. Proceeding in this manner, we find a sequence of v as in (77) which converges to an element of F.

66. We shall prove that the element just obtained—call it v_0—belongs to $\mathfrak{M}$.

Let v_0 not belong to $\mathfrak{M}$, and let it annul some form—call it B—of (59). We shall produce the contradiction that F has a strong sum.

We consider the form C_0 of (60) which is secured for B. The solutions of B which belong to $\mathfrak{M}$ annul C_0. Let $\Sigma_1, \cdots, \Sigma_s$ be a decomposition of the system B, C_0 into closed essential irreducible systems. According to § 31, each Σ_i has a basic set

$$U_i, V_i \tag{79}$$

introducing u and v in succession, with U_i algebraically irreducible and of order at most unity in u, and with V_i of order zero in v.

Let us consider those Σ_i of which $u = v = 0$ is a solution.[19] We say that, for at least one of them, U_i is of order unity. Let this be false. If a U_i is of order zero and vanishes for $u = 0$, that U_i, being algebraically irreducible, must be u multiplied by a function of x. Each Σ_i of which $u = v = 0$ is not a solution contains a form $1 + A$ where A vanishes for $u = v = 0$. We conclude that the system B, C_0 is held by a form M given by

$$M = u(1 + H) \tag{80}$$

---

[19] Because v_0 annuls B, $u = v = 0$ annuls B.

where H vanishes for $u = v = 0$. Thus M holds the system obtained by adjoining B to Φ, so that some M^g is a linear combination of B, the forms of Φ and their derivatives.

We now consider the sequence of u-expansions which converges to v_0. We substitute successively these expansions for v, and the associated expansions for u_1, into the linear expression just found for M^g. Any form of Φ, and any derivative of such a form, will yield a sequence of expansions which converges to zero. Also, for a sequence of expansions converging to v_0, and for the related u_1, B and its derivatives yield sequences which converge to zero. Thus M^g yields a sequence converging to zero. This is impossible because u^g is a term of lowest degree for M^g.[20]

Let then Σ_1 admit $u = v = 0$ as a solution and let U_1 be of order unity. We consider the equation

$$U_1 = 0 \tag{81}$$

as an algebraic equation for u_1. For x in a suitable area $\mathfrak{A}_1$, and for u small, the solutions of (81) can be expressed as series of ascending fractional powers of u with coefficients analytic in $\mathfrak{A}_1$. For x in $\mathfrak{A}_1$ and for u small but not zero, these series give finite values for u_1 which, for given values of x and u, are the only numerical solutions of (81) for u_1. None of these series annuls the initial of V_1 identically in x and u. When any one of the series is substituted for u_1 in V_1, the equation $V_1 = 0$ determines v as any one of a finite number of series of ascending fractional powers of u with coefficients which will be analytic in $\mathfrak{A}_1$ if $\mathfrak{A}_1$ is shrunk appropriately.

We obtain thus a certain number of related pairs of series for u_1 and v. For x in $\mathfrak{A}_1$ and for u small but not zero, such a pair of series gives finite values of u_1 and v, and the pairs of values of u_1 and v from all of the series will be the only numerical solutions of $U_1 = 0$, $V_1 = 0$, for given x and u.

Because u does not hold Σ_1, the solution $u = v = 0$ of Σ_1 is approximable by solutions $\bar{u}$, $\bar{v}$ with $\bar{u} \neq 0$. There are points b in $\mathfrak{A}_1$ for which we can get $\bar{u}$ and $\bar{v}$ which, with any finite number of their derivatives, are as small as one pleases at b, while $\bar{u}(b) \neq 0$. There must be some one related pair of series for u_1 and v which, for some sequence of $\bar{u}$ approximating more and more closely to $u = 0$, give $\bar{u}_1$ and $\bar{v}$ when u is replaced by $\bar{u}$. According to *A. D. E.*, § 89, the lowest exponent of u in the series for u_1 is at least unity. Because $\bar{u}(b)$ and $\bar{v}(b)$ can be made arbitrarily small, the series for $\bar{v}$ must contain only positive exponents of u. Thus there is a pair of series

---

[20] It is unnecessary to use strong characteristics here. In *S. S.*, § 81, the characteristic of (159) is not less than $\alpha + \beta$, even if α and β are not strong.

$$u_1 = \alpha_0 u + \alpha_1 u^{1+1/r} + \cdots, \tag{82}$$

$$v = \beta u^{1/r} + \cdots, \tag{83}$$

which annul U_1 and V_1. The remainder of C_0 with respect to the ascending set U_1, V_1 is zero. Because the initial of V_1 does not vanish for (82), C_0 is annulled by (82), (83). Similarly, B is annulled by (83).

We now examine (60).

Let the series in (82) and (83) be denoted by $\tilde{u}_1$ and $\tilde{v}$. Let C_0 be expanded in powers of $u_1 - \tilde{u}_1$ and $v - \tilde{v}$, with u-expansions for coefficients. The expression for C_0 will contain no term free of $u_1 - \tilde{u}_1$ and $v - \tilde{v}$. Because C_0 does not vanish identically in x, u, u_1 for $v = \tilde{v}$, the expression contains a term free of $v - \tilde{v}$. Let

$$C_0 = \gamma_1 (u_1 - \tilde{u}_1)^{g_1} + \cdots + \gamma_t (u_1 - \tilde{u}_1)^{g_t} + \cdots \tag{84}$$

where the γ are u-expansions and the g increasing positive integers. The terms which follow $\gamma_t (u_1 - \tilde{u}_1)^{g_t}$ in (84) involve $v - \tilde{v}$.

In (60), keeping C_0 in its original form as a polynomial in v, u_1, u, we make the substitutions

$$\begin{aligned} v &= \tilde{v} + w \\ v_1 &= \frac{\partial \tilde{v}}{\partial u} + \frac{\partial \tilde{v}}{\partial x} u_1 + w_1 \end{aligned} \tag{85}$$

with w an indeterminate and w_1 its derivative. Then each term in the second member of (60) goes over into a polynomial in w and w_1. The coefficients in these polynomials are polynomials in u_1 whose coefficients are series of rational powers of u. B^p will give a polynomial in w in which the least exponent of w is p. The power products $B^{p_i} B'^{q_i}$ will produce polynomials in w and w_1 with terms all of degree at least $p + 1$.

Let the coefficient of w^p in the polynomial yielded by B^p be a u-expansion in which the lowest exponent of u is h. If we replace w by u^ρ, with ρ a positive integer greater than h, we see that the replacement of v by $\tilde{v} + u^\rho$ in B^p produces a u-expansion in which the least exponent of u is $h + p\rho$.

In (84), let the sum

$$\gamma_1 (u_1 - \tilde{u}_1)^{g_1} + \cdots + \gamma_t (u - \tilde{u}_1)^{g_t} \tag{86}$$

be written as an infinite sum Σ of power products in u_1 and u. Because (86) equals $u_1 - \tilde{u}_1$ multiplied by a sum analogous to Σ, and because the terms in $\tilde{u}_1$ are all of degree at least unity, the terms of lowest degree in Σ must involve

u_1. Let k be the degree of the terms of lowest degree in Σ. Let ρ above exceed k as well as h. Then the substitution $v = \tilde{v} + w^\rho$ into C_0B^p produces a set of terms in which the terms of lowest degree are of degree $k + h + p\rho$ and involve u_1.

Now the substitution (85) with $w = u^\rho$, practiced on any term after the first in the second member of (60), yields terms in u and u_1 of degree at least $(p + 1)\rho$. Let $\rho > h + k$. Then the terms of lowest degree coming from the second member of (60) will involve u_1.

Because S in (60) is free of u_1, v_1, it must be that the substitution (85) with $w = u^\rho$ reduces F to a sum of power products in u and u_1 in which the terms of lowest degree involve u_1. If, then, v' is a sufficiently long segment of $\tilde{v}$, the replacement of v in F by $v' + u^\rho$ will produce a sum (37) in which u_1 figures in the terms of lowest degree. Thus $v' + w^\rho$ is a strong sum for F.

Thus v_0 is an element of F in $\mathfrak{M}$. The necessity of the fulfillment of at least one of the conditions of § 54 is established.

SUMMARY OF TEST.

67. Let us summarize the method for testing whether $u = v = 0$ is contained in $\mathfrak{M}$, the general solution of F. We recall that F is algebraically irreducible, that the maximum of the orders of F in u and v is unity and that $u = v = 0$ is a solution of F.

One first secures the decomposition (31). If there are no B_i as in (31) or no such B_i which are annulled by $u = v = 0$, then $u = v = 0$ is in $\mathfrak{M}$. In what follows, we assume that B_i exist, in (31), which vanish for $u = v = 0$.

One tests now as in Part V to see whether the manifold of the form u is contained in $\mathfrak{M}$. An affimative answer means that $u = v = 0$ is in $\mathfrak{M}$. In what follows, we assume that $\mathfrak{M}$ does not contain the manifold of u.

We compare ζ_t of § 37 with m of § 50, extending the definition of m so as to have $m = 0$ if there are no B_i as in (59). If $\zeta_t \neq m$, $u = v = 0$ is in $\mathfrak{M}$. If $\zeta_t = m$, $u = v = 0$ is or is not in $\mathfrak{M}$ according as the index of F is ∞ or ζ_t; the index is determined as in § 51.[21]

ONE-PARAMETER FAMILIES.

68. In § 31 it was stated that, if U in (32) is of order unity, the manifold of Σ is contained in $\mathfrak{M}$. We show now how this is proved.

Let $u = \phi$, $v = \psi$ be any solution of Σ.

---

[21] If $\zeta_t = 0$, the index is 0. Otherwise the single point of which $\mathfrak{P}$ is composed would be a b-point. This would imply that $\mathfrak{M}$ contains the manifold of u.

Following § 66, we prove that U, V, B, C_0 are each annulled by a pair of series

$$
\begin{aligned}
u_1 &= \phi_1 + \alpha_0(u - \phi) + \alpha_1(u - \phi)^{1+1/r} + \cdots, \\
v &= \psi + \beta(u - \phi)^{1/r} + \cdots,
\end{aligned}
\tag{87}
$$

with ϕ_1 the derivative of ϕ. We then use (60) to show that the substitution

$$
v = \bar{v} + (u - \phi)^{\rho}
\tag{88}
$$

with $\bar{v}$ the series for v in (87) and ρ a sufficiently large integer, reduces F to a sum of power products in $u - \phi$ and $u_1 - \phi_1$ with $u_1 - \phi_1$ present in the products of lowest degree. The argument of § 53 is then used to show that the solution $u = \phi$, $v = \psi$ of F can be approximated by solutions in $\mathfrak{M}$. Thus $u = \phi$, $v = \psi$ is in $\mathfrak{M}$.

EXAMPLES.

69. *Example* 1. Let

$$
A = u(uv_1 + uu_1 - 2u_1v) - (v - u)^2.
$$

Let F be the form, algebraically irreducible in the field of all rational functions of x, defined by

$$
uF = vA^2 - \prod_{j=0}^{4}\left(v - u + \frac{u^2}{x + j}\right).
$$

Then $\mathfrak{P}$ has no b-point, $\zeta_t = 4$, $m = 5$. $\mathfrak{M}$ contains $u = v = 0$.

Example 2. Let

$$
A = uv_1 + uu_1 - 2u_1v.
$$

Let

$$
F = A^2 + \prod_{j=0}^{2}(v - u + ju^2).
$$

Here $\mathfrak{P}$ has no b-points and $\zeta_t = m = 3$. Carrying out the substitution $v = u + v'$, we find that $\mathfrak{P}'$ has a b-point, so that $u = v = 0$ is in $\mathfrak{M}$.

Example 3. Let $A = v^2 - u^3$ and let A_1 be the derivative of A. Let

$$
F = A_1{}^2 - A.
$$

Then $\mathfrak{P}$ has no b-point, and $\zeta_t = m = 2$. By § 51 we see that the index is 2, so that $u = v = 0$ is not in $\mathfrak{M}$.

Example 4. The similarity of the preceding examples to examples of *S. S.* might lead one to ask whether the problem of the present paper cannot be reduced to that of *S. S.* by replacing v by a form of the first order in u, for instance, by u_1. Let $F = u + v_1^2$. Then $\mathfrak{P}$ has a b-point so that $u = v = 0$ is in $\mathfrak{M}$. If we substitute for v a form of the first order in u which vanishes for $u = 0$, F becomes a form in u for which $u = 0$ is an essential manifold.

Example 5. Let

$$F = v_1^2(v_1 - u_1) + v(v - u).$$

The manifold of F decomposes into $\mathfrak{M}$ and the manifold of v. $\mathfrak{P}$ has no b-point, $\zeta_t = 2$, $m = 1$. We observe that zero and u are elements. By § 41, if F had a strong sum, the first term of the strong sum would be u. If we put, in F,

$$v = u + \phi_2 u^{\rho_2} + \cdots + \phi_k u^{\rho_k}$$

with $\phi_2 \neq 0$, $\rho_2 > 1$, the term $v(v - u)$ in F produces a sum of powers of u in which the least exponent is $\rho_2 + 1$. The first term of F produces power products in u and u_1 of degree no less than $\rho_2 + 2$. There is consequently no strong sum, and the index is 2.

One might ask whether, in the case in which F has a finite index and there are elements in $\mathfrak{M}$, the elements in $\mathfrak{M}$ have to be algebraic with respect to u. At least in the case in which F has constant coefficients, the answer is affirmative. The general question should be interesting to examine.

Example 6. Let $F = v_1^2 + (u_1 + v)v$. The manifold of F consists of $\mathfrak{M}$ and the manifold of v. The two irreducible manifolds have in common the one-parameter family $u = c$, $v = 0$.

COLUMBIA UNIVERSITY.

THE ANALYSIS OF THE DIRECT PRODUCT OF IRREDUCIBLE REPRESENTATIONS OF THE SYMMETRIC GROUPS.*

By F. D. Murnaghan.

Let Γ and Γ' denote irreducible representations of the symmetric groups on m and n letters, respectively, with characteristics ϕ and ϕ'. Their direct product $\Gamma \cdot \Gamma'$ is a representation, in general reducible, of the symmetric group on $m+n$ letters with characteristic $\phi\phi'$. We have given in a previous paper (1) the analysis of $\Gamma \cdot \Gamma'$ into its irreducible components for all representations Γ, Γ' for which $m+n \leq 9$. In the present paper we present a refinement of the method used in (1) which makes the computations very much easier and add the table giving the analyses of the products $\Gamma \cdot \Gamma'$ for which $m+n=10$. The simple characteristics ϕ of the symmetric group have been termed Schur-Functions ($=S$-Functions) and the problem under consideration has been treated recently under the title "Multiplication of S-Functions" by Littlewood and Richardson (2) who have proposed a scheme involving the construction of various tableaux (based on the partitions of n with which the irreducible representations of the symmetric group on n letters are associated). They have, however, been unable to present a proof, in the general case, of the theorem on which their proposed scheme rests; and, unfortunately, the example they give to illustrate the operation of their method (namely, the direct product of the irreducible representation $\Gamma = D(4, 3, 1)$, of dimension 70, of the symmetric group on 8 letters by the irreducible representation $\Gamma' = D(2^2, 1)$, of dimension 5, of the symmetric group on 5 letters) has the result incorrectly printed. The tableaux they give furnish the correct result and so the error in the final result must be ascribed to the printer or to an oversight in reading off the representations associated with the various tableaux (of which there are 34). That this error was not immediately detected will not appear surprising when we remark that the direct product being analysed is a representation of dimension 450, 450 of the symmetric group on 13 letters. The method we give in the present paper makes it easy to read off the analysis of this representation in less than five minutes. We furnish rules, with illustrative examples, which make the analysis of $\Gamma \cdot \Gamma'$ simple if $m+n \leq 16$. These rules depend for their construction on the tables, given in our paper (1) for $n \leq 9$, which furnish the analysis of the reducible representations $\Delta(\lambda)$ of the symmetric group on n letters; and on tables, reciprocal to these, which express

---

* Received October 6, 1937.

each simple characteristic of the symmetric group on n letters as a linear combination of the characters of the reducible representations $\Delta(\lambda)$ of this group. These tables were originally furnished by Kostka ((**3**), (**4**), (**5**)) in connection with his fundamental researches on symmetric functions; his method for constructing them is, however, quite unnecessarily complicated and it is not surprising that the usefulness of his tables for the present problem has hitherto escaped attention. As his tables may not be easily accessible to those working in nuclear physics (for whom, principally, this paper is written) we give the necessary "reciprocal tables" for $n \leq 9$. The calculation of these tables is very simple, involving not more than a few hours work; in fact the simplest construction of the tables of our previous paper, which furnished the analysis of $\Delta(\lambda)$, appears to be the following. First construct the tables of the present paper; each table is a triangular matrix with diagonal elements unity and, hence, having its determinant unity. The tables furnishing the analysis of $\Delta(\lambda)$ are obtained by taking the reciprocals of these triangular matrices; a procedure involving merely a recurrent transposition of terms from one side of a system of linear equations to the other.

Since each simple characteristic ϕ of the symmetric group on n letters is the character of a rational, homogeneous, irreducible representation, of degree n, of the full linear group (= group of all non-singular linear homogeneous transformations in p variables) (**6**) the results of the present paper furnish the analysis of the Kronecker product of any two such rational, homogeneous, irreducible representations of the full linear group.

It is pleasant to conclude this introduction by remarking that our interest in the problem has been greatly stimulated by the queries of our friends Professors J. A. Wheeler and E. Wigner whose fundamental researches in nuclear physics require the analysis of $\Gamma \cdot \Gamma'$ given here.

1. Notations and outline of the general method.[1] Let

$$(\epsilon) = (\epsilon_1, \epsilon_2, \cdots, \epsilon_n), \qquad \epsilon_1 \geq \epsilon_2 \geq \cdots \geq \epsilon_n \geq 0$$

be any partition of n and let $\phi_{(\epsilon)}(s)$ be the simple characteristic of the symmetric group on n letters which is associated with the partition (ϵ); similarly let

$$(\nu) = (\nu_1, \nu_2, \cdots, \nu_m), \qquad \nu_1 \geq \nu_2 \geq \cdots \geq \nu_m \geq 0$$

be a partition of m and let $\phi_{(\nu)}(s)$ be the corresponding simple characteristic of the symmetric group on m letters. Then the product $\phi_{(\epsilon)}(s)\phi_{(\nu)}(s)$ is a linear combination of the simple characteristics $\phi_{(\alpha)}(\boldsymbol{s})$ of the symmetric group on $n+m$ letters. We have given (p. 480) the following rule for determining the coefficients of this linear combination. Regarding $(\boldsymbol{s}) = (s_1, s_2, \cdots)$ as the power sums of $n+m$ variables

---

[1] We shall assume the reader familiar with the methods and notations of our previous paper (**1**) and shall refer, where convenient, to this paper by merely giving the page number.

$$(\boldsymbol{z}) = (z_1, z_2, \cdots, z_{n+m}) \,:\, s_j = z_1^j + \cdots + z_{n+m}^j, \qquad (j = 1, 2, \cdots)$$

$\phi_{(\epsilon)}(\boldsymbol{s})$ is a linear combination $\sum_{(\pi)} c_{(\pi)} T_{(\pi)}(\boldsymbol{z})$ of the symmetric functions $T_{(\pi)}(\boldsymbol{z}) = \Sigma z_1^{\pi_1} \cdots z_n^{\pi_n}$ of degree n in the $n+m$ variables $(\boldsymbol{z})$ (the summation in the expression for $\phi_{(\epsilon)}(\boldsymbol{s})$ being over all partitions (π) of n). Then those characteristics $\phi_{(\alpha)}(\boldsymbol{s})$ of the symmetric group on $n+m$ letters occur for which any member of $[(\alpha) - (\pi)]$ is the partition (ν) of m. It is understood that $[(\alpha) - (\pi)]$ means the aggregate of partitions of m obtained by subtracting (π), in all possible arrangements, from each group of n from the $n+m$ letters (α); and further that any disordered arrangement in the set $[(\alpha) - (\pi)]$ is restored to the normal non-increasing arrangement in the manner described in (1) (p. 461). An equivalent statement of this result is clearly the following: add n zeros to (ν) so that it appears as a partition of m with $n+m$ elements $(\nu) = (\nu_1, \nu_2, \cdots, \nu_m, \overbrace{0, 0, \cdots, 0}^{n \text{ zeros}})$ and add (π) in all possible arrangements to each set of n from the $n+m$ numbers (ν). The resulting partitions of $n+m$ are rearranged, if disordered, according to the rule referred to and those which do not vanish will appear in the product $\phi_{(\epsilon)}(\boldsymbol{s})\phi_{(\nu)}(\boldsymbol{s})$ with the coefficient $\pm c_{(\pi)}$; the $+$ sign being used if an even, and the $-$ sign if an odd, number of inversions are necessary to bring the disordered partition of $n+m$ into its natural order.

It is clear that a slavish adherence to the rule just given would prove tedious and it would in fact be indicative of a lack of intelligence. For if the expression $\sum_{(\pi)} c_{(\pi)} T_{\pi}(\boldsymbol{z})$ be written as a polynomial in z_1 the various coefficients are symmetric functions in the $n+m-1$ variables $z_2, \cdots, z_{n+m}$ the coefficient of z_1^p being of degree $n-p$. Each such symmetric function of degree $n-p$ may be expressed as a linear combination of the simple characteristics of the symmetric group on $n-p$ letters; let $a_{(\lambda')}\phi_{(\lambda')}(\boldsymbol{s}')$ be a term of this linear combination where the $\boldsymbol{s}'$ are the power sums of the $n+m-1$ variables $z_2, \cdots, z_{n+m}$. Then amongst the terms of the desired product $\phi_{(\epsilon)}(\boldsymbol{s})\phi_{(\nu)}(\boldsymbol{s})$ will occur terms $a_{(\lambda')}\phi_{(\alpha)}(\boldsymbol{s})$ where (α) is obtained from any term $\phi_{(\beta)}(\boldsymbol{s}')$ occurring in the product $\phi_{(\nu')}(\boldsymbol{s}')\phi_{(\lambda')}(\boldsymbol{s}')$ (where $(\nu') = (\nu_2, \cdots, \nu_m)$) by prefixing an element $\nu_1 + p$. We hope the reader will not be confused by the number of words necessary to state the rule which is really simpler to apply than to describe; a few elementary illustrations will make it clear.

The most elementary (indeed trivial) instance of our problem occurs when $n = 1$. Here there is only one partition $(\pi) = (1)$ and only one simple characteristic $\phi_{(\epsilon)}(\boldsymbol{s}) = \phi_{(1)}(\boldsymbol{s}) = s_1 = T_{(1)}(\boldsymbol{z})$ and the original form of statement of our rule is immediately applicable. We write (ν) in the form $(\nu_1, \cdots, \nu_m, 0)$ and add, in turn, unity to each element obtaining the result (p. 480)

$$\{1\} \cdot \{\nu_1, \cdots, \nu_m\} = \{\nu_1 + 1, \cdots, \nu_m\} + \cdots \\ + \{\nu_1, \cdots, \nu_m + 1\} + \{\nu_1, \cdots, \nu_m, 1\}$$

where we use the notation $\{\nu_1, \cdots, \nu_m\}$ to denote $\phi_{(\nu)}(\boldsymbol{s})$. If $\nu_k > 0$, $\nu_{k+1} = \nu_{k+2} = \cdots = \nu_m = 0$ we omit the terms on the right in which there appears a unit preceded by a zero (p. 461) and our result appears in the simpler form

$$\{1\} \cdot \{\nu_1, \cdots, \nu_k\} = \{\nu_1 + 1, \cdots, \nu_k\} + \{\nu_1, \cdots, \nu_k + 1\} + \{\nu_1, \cdots, \nu_k, 1\}.$$

The next case, namely multiplication by $\{2\}$, is not quite so trivial and enables us to glimpse the advantages of the modification, given above, of the original form of statement. We first remark that since $p_0(\boldsymbol{z}) = 1$ (p. 449) $\{0, 0, 0, \cdots\} = 1$ so that $\{0, 0, 0, \cdots\} \cdot \{\nu_1, \cdots, \nu_k\} = \{\nu_1, \cdots, \nu_k\}$. Now $\phi_{(2)}(\boldsymbol{s}) = p_2(\boldsymbol{z}) = T_{(2)}(\boldsymbol{z}) + T_{(1^2)}(\boldsymbol{z})$ and writing this as a polynomial (of the second degree) in z_1 the terms independent of $z_1 = \phi_2(\boldsymbol{s}')$; the coefficient of $z_1 = \phi_1(\boldsymbol{s}')$ whilst the coefficient of $z_1{}^2$ is unity. Hence

$$\{2\} \cdot \{\nu_1\} = \{2\} \cdot \{\nu_1, 0, 0, \cdots, 0\} = \{\nu_1, 2\} + \{\nu_1 + 1, 1\} + \{\nu_1 + 2\}.$$

From this we deduce

$$\{2\} \cdot \{\nu_1, \nu_2\} = \{\nu_1, \nu_2, 2\} + \{\nu_1, \nu_2 + 1, 1\} + \{\nu_1, \nu_2 + 2\} \\ + \{\nu_1 + 1, \nu_2, 1\} + \{\nu_1 + 1, \nu_2 + 1\} + \{\nu_1 + 2, \nu_2\}$$

and so on in general; the final result being that $\{2\} \cdot \{\nu_1, \nu_2, \cdots, \nu_j\}$ is obtained by writing $\{\nu_1, \nu_2, \cdots, \nu_j\}$ in the form $\{\nu_1, \nu_2, \cdots, \nu_j, 0\}$ and then adding 2 and $(1, 1)$ in all possible ways (p. 480).

Multiplication by $\{1^2\}$ is even simpler since $\phi_{(1^2)}(\boldsymbol{s}) = \sigma_2(\boldsymbol{z}) = T_{(1^2)}(\boldsymbol{z})$. Writing this as a polynomial (of the first degree) in z_1 the terms independent of $z_1 = \phi_{(1^2)}(\boldsymbol{s}')$ whilst the coefficient of $z_1 = \phi_1(s')$. Hence

$$\{1^2\} \cdot \{\nu_1\} = \{\nu_1, 1, 1\} + \{\nu_1 + 1, 1\}$$

and so on, in general, the final result being that $\{1^2\} \cdot \{\nu_1, \nu_2, \cdots, \nu_j\}$ is obtained by writing $\{\nu_1, \cdots, \nu_j\}$ in the form $\{\nu_1, \cdots, \nu_j, 0, 0\}$ and then adding $\{1, 1\}$ in all possible ways (p. 481).

We hope that it is clear from the preceding paragraphs that the essential steps in the calculation of $\phi_{(\epsilon)}(\boldsymbol{s}) \cdot \phi_{(\nu)}(\boldsymbol{s})$ are the following:

(1) First express $\phi_{(\epsilon)}(s)$ in the form $\sum_{(\pi)} c_{(\pi)} T_{(\pi)}(z)$ and write it as a polynomial in z_1. This step in the calculation has already been done for us (as far as $n = 11$) in the tables of Kostka referred to in the introduction; we shall discuss it in the following paragraph.

(2) Next express the coefficients of the various powers of z_1 as linear combinations of the simple characteristics of the appropriate symmetric groups (the coefficient of z_1^p being expressible in terms of the simple characteristics of the symmetric group on $n - p$ letters). This step has also been done for us in Kostka's tables; we give below tables furnishing the necessary coefficients (as far as $n = 9$).

When these steps have been performed the product $\phi_{(\epsilon)}(s) \cdot \phi_{(\nu)}(s)$ can be at once written down if the products $\phi_{(\lambda)}(s) \cdot \phi_{(\nu')}(s)$ are known where $(\nu') = (\nu_2, \cdots, \nu_j)$ and (λ) is a partition of n, or less than n, letters. Thus the desired products are obtainable without difficulty by a recurrence method. Before proceeding to a description of the methods by which steps (1) and (2) are carried out in general it is probably desirable to illustrate the procedure by another simple example.

Example. $\{2, 1\} \cdot \{2, 1\}$.

We shall see below that $\{2, 1\} = T_{(2,1)}(z) + 2T_{(1^3)}(z)$. The terms free of z_1 in this second degree polynomial $= \phi_{(2,1)}(s')$ and so the characteristics beginning with 2 in the desired product are found by prefixing a 2 to the product $\{1\} \cdot \{2, 1\}$ i. e. $\{3, 1\} + \{2^2\} + \{2, 1^2\}$. Since $\{2, 3, 1\} = 0$ (1, p. 461) they are, accordingly, $\{2^3\} + \{2^2, 1^2\}$. The coefficient of z_1 in our second degree polynomial is $\Sigma z_2{}^2 + 2\Sigma z_2 z_3 = \phi_2(s') + \phi_{(1^2)}(s')$ and so the characteristics beginning with 3 in the desired product are found by prefixing a 3 to the product $\{1\} \cdot [\{2\} + \{1^2\}]$ i. e. $\{3\} + 2\{2, 1\} + \{1^3\}$. They are, accordingly, $\{3^2\} + 2\{3, 2, 1\} + \{3, 1^3\}$. The coefficient of $z_1{}^2$ is $\phi_1(s')$ and so the characteristics beginning with 4 in the desired product are found by prefixing a 4 to the product $\{1\} \cdot \{1\}$ i. e. $\{2\} + \{1^2\}$. They are, accordingly, $\{4, 2\} + \{4, 1^2\}$. Hence the desired product is

$$\{4, 2\} + \{4, 1^2\} + \{3^2\} + 2\{3, 2, 1\} + \{3, 1^3\} + \{2^3\} + \{2^2, 1^2\}$$

2. Expression of the simple characteristics $\phi_{(\epsilon)}(s)$ of the symmetric group on n letters in terms of the symmetric functions $T_{(\pi)}(z)$. This often treated problem is so simple that it can be discussed *ab initio* in a few lines. $\phi_{(\epsilon)}(s)$ may be presented as a determinant of order n whose diagonal elements are $p_{\epsilon_1}(z), p_{\epsilon_2}(z), \cdots, p_{\epsilon_n}(z)$, the non-diagonal elements being obtained by methodically increasing (decreasing) by unity the labels attached to the $p(z)$ as we move from each column to its neighbor on the right (left). An equivalent statement of this fact is the following: let ξ_j be an operator which decreases by unity the value of the label carrying the subscript j $(j = 1, 2, \cdots, n)$ and let $e_1 = \epsilon_1 + (n-1)$, $e_2 = \epsilon_2 + (n-2), \cdots, e_{n-1} = \epsilon_{n-1} + 1$, $e_n = \epsilon_n$ (so that $e_1 > e_2 > \cdots > e_n \geqq 0$). Then

$$\phi_{(\epsilon)}(s) = \begin{vmatrix} \xi_1^{n-1} & \cdots & \xi_n^{n-1} \\ \xi_1^{n-2} & \cdots & \xi_n^{n-2} \\ \cdot & \cdots & \cdot \\ 1 & \cdots & 1 \end{vmatrix} p_{e_1} p_{e_2} \cdots p_{e_n}.$$

If, therefore, we denote by $K_{(\lambda)}(z)$ the characteristic $p_{\lambda_1} \cdots p_{\lambda_n}$ of the reducible representation $\Delta(\lambda)$ of the symmetric group on n letters we see that $\phi_{(\epsilon)}(s)$ is a linear combination of the various compound characteristics $K_{(\lambda)}(z)$; only those characteristics $K_{(\lambda)}(z)$ occurring for which (λ) increased by the set $(n-1, n-2, \cdots, 1, 0)$, arranged in any order, gives the set $(e) = (e_1, \cdots, e_n)$ arranged in any order. The sign attached to $K_{(\lambda)}(z)$ is plus (minus) if the arrangement of the set $(n-1, n-2, \cdots, 1, 0)$ which is added to (λ) to give (e) is even (odd). E. g., let $(\epsilon) = (3, 2, 1)$ so that $(e) = (5, 3, 1)$; to find the $K_{(\lambda)}(z)$ which occur in the expression for $\phi_{(3,2,1)}(s)$ we subtract $(2, 1, 0)$ in all possible orders from $(5, 3, 1)$ obtaining the six terms $(3, 2, 1), -(3, 3, 0)$, $(5, 1, 0)$, $-(5, 2, -1)$, $(4, 3, -1)$, $-(4, 1, 1)$. Of these the fourth and fifth vanish since they contain a negative element (each $p_j(z)$ for which $j < 0$ being zero) and so $\phi_{(3,2,1)} = K_{(3,2,1)} - K_{(3^2)} - K_{(4,1^2)} + K_{(5,1)}$. The same rule holds if we wish to find the expression for any symmetric function $T_{(\lambda)}(z) = \Sigma z_1^{\lambda_1} \cdots z_n^{\lambda_n}$ of degree n in terms of the simple characteristics $\phi_{(\epsilon)}(s)$. In fact $\phi_{(\epsilon)}(s)$ may be presented as the ratio

$$A(e_1, \cdots, e_n) : A(n-1, \cdots, 1, 0)$$

where $A(v_1, \cdots, v_n)$ is the n-rowed determinant of which the elements in the j-th row are the v^j-th powers of the indeterminates $(z_1, \cdots, z_n)$ $(j = 1, \cdots, n)$ (p. 459). On multiplying $T_{(\lambda)}(z)$ by the Vandermonde determinant $\Delta(z) = A(n-1, \cdots, 1, 0)$ we obtain a collection of determinants $A(v_1, \cdots, v_n)$ and on dividing through by $A(n-1, \cdots, 1, 0)$ we see that $T_{(\lambda)}(z)$ is expressible as a linear combination of the simple characteristics $\phi_{(\epsilon)}(s)$; only those simple characteristics $\phi_{(\epsilon)}(s)$ occurring for which (ϵ) increased by the set $(n-1, \cdots, 1, 0)$ i. e. (e) is the same as one of the sets (λ) + any arrangement of the set $(n-1, \cdots, 1, 0)$. The sign attached to $\phi_{(\epsilon)}(s)$ is plus (minus) if the arrangement of the set $(n-1, \cdots, 1, 0)$ which is added to (λ) to give (e) is even (odd). Hence the coefficient of $K_{(\lambda)}(z)$ in the development of $\phi_{(\epsilon)}(s)$ is the same as the coefficient of $\phi_{(\epsilon)}(s)$ in the development of $T_{(\lambda)}(z)$ (a result apparently first observed by Kostka). The second step in our general procedure is therefore easy provided we know the expansions of the simple characteristics $\phi_{(\epsilon)}(s)$ in terms of the compound characteristics $K_{(\lambda)}(z)$. For partitions (ϵ) containing not more than three non-zero elements these expansions are trivial since they involve nothing more than the expansion of a determinant of order three or less. If (ϵ) has only one non-zero element $\phi_{(\epsilon)}(s)$ appears as a determinant of one row and $\phi_{(\epsilon)}(s) = K_{(\epsilon)}(z)$. E. g., $\phi_{(4)}(s) = K_{(4)}(z)$. For partitions with two non-zero elements we have to expand a two-row determinant; E. g.,

$$\phi_{(4,2)}(s) = \begin{vmatrix} p_4(z) & p_5(z) \\ p_1(z) & p_2(z) \end{vmatrix} = p_4(z)p_2(z) - p_5(z)p_1(z)$$

so that $\phi_{(4,2)}(s) = K_{(4,2)}(z) - K_{(5,1)}(z)$. For three non-zero element partitions we proceed similarly; E. g.,

$$\phi_{(3,2^2)}(s) = \begin{vmatrix} p_3(z) & p_4(z) & p_5(z) \\ p_1(z) & p_2(z) & p_3(z) \\ p_0(z) & p_1(z) & p_2(z) \end{vmatrix} \begin{aligned} &= p_3(z)p_2{}^2(z) - p_3{}^2(z)p_1(z) \\ &\quad - p_4(z)p_2(z)p_1(z) + p_5(z)p_1{}^2(z) \\ &\quad + p_4(z)p_3(z) - p_5(z)p_2(z) \end{aligned}$$

and so

$$\begin{aligned} \phi_{(3,2^2)}(s) = K_{(3,2^2)}(z) \\ - K_{(3^2,1)}(z) - K_{(4,2,1)}(z) + K_{(4,3)}(z) + K_{(5,1^2)}(z) - K_{(5,2)}(z). \end{aligned}$$

For partitions containing four or more elements it is convenient to expand $\phi_{(\epsilon)}(s)$ in terms of the first column the cofactors of the elements in this column being simple characteristics corresponding to three element partitions. E. g., suppose we wish $\phi_{(5,2,1^2)}(s)$: from the expression

$$\phi_{(5,2,1^2)}(s) = \begin{vmatrix} p_5 & p_6 & p_7 & p_8 \\ p_1 & p_2 & p_3 & p_4 \\ 0 & p_0 & p_1 & p_2 \\ 0 & 0 & p_0 & p_1 \end{vmatrix}$$

we read $\phi_{(5,2,1^2)}(s) = p_5\phi_{(2,1^2)}(s) - p_1\phi_{(3,1^2)}(s)$ and from the supposed known analyses of $\phi_{(2,1^2)}(s)$ and $\phi_{(6,1^2)}(s)$ we find (since $p_5K_{(2,1^2)}(z) = K_{(5,2,1^2)}(z)$ and so on)

$$\begin{aligned} \phi_{(5,2,1^2)}(s) = - K_{(8,1)}(z) + K_{(7,1^2)}(z) + K_{(6,2,1)}(z) - K_{(6,1^3)}(z) \\ + K_{(5,4)}(z) - K_{(5,3,1)}(z) - K_{(5,2^2)}(z) + K(_{(5,2,1^2)}(z). \end{aligned}$$

We give below tables furnishing the expressions for the simple characteristics $\phi_{(\epsilon)}(s)$ in terms of the compound characteristics $K_{(\lambda)}(z)$ for all values of n up to 9 inclusive. The simple characteristics are written down the left-hand side of each table (the partitions (ϵ) of n being arranged in dictionary order) and the compound characteristics are written across the top of each table (the partitions (λ) of n being also arranged in dictionary order). The matrix connecting the $\phi_{(\epsilon)}(s)$ with the $K_{(\lambda)}(z)$ is triangular with diagonal elements all unity and, hence, is of determinant unity; the triangular reciprocal matrix expresses the compound characteristics $K_{(\lambda)}(z)$ in terms of the simple characteristics $\phi_{(\epsilon)}(s)$. These reciprocal matrices have been given, for $2 \leq n \leq 9$, in our paper **(1)** (pp. 475-477). From Kostka's theorem stating that the coefficient of $K_{(\lambda)}(z)$ in the development of $\phi_{(\epsilon)}(s)$ is the same as the coefficient of $\phi_{(\epsilon)}(s)$ in the development of $T_{(\lambda)}(z)$ we see that the coefficients of this latter development are found in the column headed by the partition (λ) in the tables which furnish the expressions for the $\phi_{(\epsilon)}(s)$ in terms of the $K_{(\lambda)}(z)$. As examples of how these tables are read we cite the following:[2]

$$\{1^2\} = -K(2) + K(1^2); \qquad T(2) = \{2\} - \{1^2\}$$
$$\{2,1\} = -K(3) + K(2,1); \qquad T(2,1) = \{2,1\} - 2\{1^3\}$$
$$\{2,1^2\} = K(4) - K(3,1) - K(2^2) + K(2,1^2);$$
$$T(2^2) = \{2^2\} - \{2,1^2\} + \{1^4\}$$
$$\{2^3\} = -K(4,2) + K(4,1^2) + K(3^2) - 2K(3,2,1) + K(2^3);$$
$$T(3,2,1) = \{3,2,1\} - 2\{3,1^3\} - 2\{2^3\} + 4\{2,1^4\} - 6\{1^6\}.$$

If we denote by $c^{(\lambda)}_{(\epsilon)}$ the coefficient of $K_{(\lambda)}(z)$ in the development of $\phi_{(\epsilon)}(s)$:

$$\phi_{(\epsilon)}(s) = \sum_{(\lambda)} c^{(\lambda)}_{(\epsilon)} K_{(\lambda)}(z)$$

Kostka's theorem finds its expression in the formula

$$T_{(\lambda)}(z) = \sum_{(\epsilon)} c^{(\lambda)}_{(\epsilon)} \phi_{(\epsilon)}(s).$$

Expressed in technical terms this says the matrix of the linear transformation from the $\phi_{(\epsilon)}(s)$ to the $T_{(\lambda)}(z)$ is the transpose of the matrix of the linear transformation from the $K_{(\lambda)}(z)$ to the $\phi_{(\epsilon)}(s)$. Since the reciprocal of the transpose of a matrix is the transpose of the reciprocal of the matrix it follows that if we write

$$K_{(\lambda)}(z) = \sum_{(\epsilon)} d^{(\epsilon)}_{(\lambda)} \phi_{(\epsilon)}(s)$$

then

$$\phi_{(\epsilon)}(s) = \sum_{(\lambda)} d^{(\epsilon)}_{(\lambda)} T_{(\lambda)}(z).$$

In other words the coefficients of the expressions for the simple characteristics $\phi_{(\epsilon)}(s)$ in terms of the symmetric functions $T_{(\lambda)}(z)$ are found in the column headed by the partition (λ) in the tables on pp. 475-477 of our paper **(1)**. As examples of how these tables are read we cite the following:

---

[2] For convenience we denote $\phi_{(\epsilon)}(s)$ by $\{\epsilon\}$; $K_{(\lambda)}(z)$ by $K(\lambda)$; $T_{(\lambda)}(z)$ by $T(\lambda)$.

$$\{2\} = T(2) + T(1^2); \qquad \{2,1\} = T(2,1) + 2T(1^3)$$

$$\{2^2\} = T(2^2) + T(2,1^2) + 2T(1^4);$$

$$\{3,1^2\} = T(3,1^2) + T(2^2,1) + 3T(2,1^3) + 6T(1^5)$$

$$\{3^2\} = T(3^2) + T(3,2,1) + T(3,1^3) + T(2^3) + 2T(2^2,1^2) + 3T(2,1^4) + 5T(1^6)$$

We have, accordingly, in the tables of this and the preceding paper the information necessary to carry out (as far as $n \leq 9$, $m \leq 9$) the steps (1) and (2) of our general method.

3. Tables furnishing the expressions for the simple characteristics $\phi_{(\epsilon)}(s)$ in terms of the compound characteristics $K_{(\lambda)}(z)$ for values of n from 2 to 9 inclusive. In the following tables the $\phi_{(\epsilon)}(s)$ are denoted by $\{\epsilon\}$ and appear down the left-hand side of the table whilst the $K_{(\lambda)}(z)$ are denoted by (λ) and appear across the top of the table. For convenience of printing, Table 8, $n = 9$, is turned around so that the bottom of the page is the left-hand side of the table and the left-hand side of the page the top of the table. The numbers to the right of the main diagonal of each table are all zero and are not written in

1. n = 2.	(2)	(1²)
{2}	1	
{1²}	−1	1

2. n = 3.	(3)	(2,1)	(1³)
{3}	1		
{2,1}	−1	1	
{1³}	1	−2	1

3. n = 4.	(4)	(3,1)	(2²)	(2,1²)	(1⁴)
{4}	1				
{3,1}	−1	1			
{2²}	0	−1	1		
{2,1²}	1	−1	−1	1	
{1⁴}	−1	2	1	−3	1

4. n = 5.	(5)	(4,1)	(3,2)	(3,1²)	(2²,1)	(2,1³)	(1⁵)
{5}	1						
{4,1}	−1	1					
{3,2}	0	−1	1				
{3,1²}	1	−1	−1	1			
{2²,1}	0	1	−1	−1	1		
{2,1³}	−1	1	2	−1	−2	1	
{1⁵}	1	−2	−2	3	3	−4	1

5. n = 6.	(6)	(5,1)	(4,2)	(4,1²)	(3²)	(3,2,1)	(3,1³)	(2³)	(2²,1²)	(2,1⁴)	(1⁶)
{6}	1										
{5,1}	−1	1									
{4,2}	0	−1	1								
{4,1²}	1	−1	−1	1							
{3²}	0	0	−1	0	1						
{3,2,1}	0	1	0	−1	−1	1					
{3,1³}	−1	1	1	−1	1	−2	1				
{2³}	0	0	−1	1	1	−2	0	1			
{2²,1²}	0	−1	1	1	0	0	−1	−1	1		
{2,1⁴}	1	−1	−2	1	−1	4	−1	1	−3	1	
{1⁶}	−1	2	2	−3	1	−6	4	−1	6	−5	1

6. n = 7.

	(7)	$(6,1)$	$(5,2)$	$(5,1^2)$	$(4,3)$	$(4,2,1)$	$(4,1^3)$	$(3^2,1)$	$(3,2^2)$	$(3,2,1^2)$	$(3,1^4)$	$(2^3,1)$	$(2^2,1^3)$	$(2,1^5)$	(1^7)
$\{7\}$	1														
$\{6,1\}$	−1	1													
$\{5,2\}$	0	−1	1												
$\{5,1^2\}$	1	−1	−1	1											
$\{4,3\}$	0	0	−1	0	1										
$\{4,2,1\}$	0	1	0	−1	−1	1									
$\{4,1^3\}$	−1	1	1	−1	1	−2	1								
$\{3^2,1\}$	0	0	1	0	−1	−1	0	1							
$\{3,2^2\}$	0	0	−1	1	1	−1	0	−1	1						
$\{3,2,1^2\}$	0	−1	0	1	1	1	−1	−1	−1	1					
$\{3,1^4\}$	1	−1	−1	1	−2	2	−1	2	1	−3	1				
$\{2^3,1\}$	0	0	1	−1	−1	0	1	2	−1	−2	0	1			
$\{2^2,1^3\}$	0	1	−1	−1	0	0	1	−1	2	1	−1	−2	1		
$\{2,1^5\}$	−1	1	2	−1	2	−4	1	−2	−3	6	−1	3	−4	1	
$\{1^7\}$	1	−2	−2	3	−2	6	−4	3	3	−12	5	−4	10	−6	1

7. n = 8.

	(8)	$(7,1)$	$(6,2)$	$(6,1^2)$	$(5,3)$	$(5,2,1)$	$(5,1^3)$	(4^2)	$(4,3,1)$	$(4,2^2)$	$(4,2,1^2)$	$(4,1^4)$	$(3^2,2)$	$(3^2,1^2)$	$(3,2^2,1)$	$(3,2,1^3)$	$(3,1^5)$	(2^4)	$(2^3,1^2)$	$(2^2,1^4)$	$(2,1^6)$	(1^8)
$\{8\}$	1																					
$\{7,1\}$	−1	1																				
$\{6,2\}$	0	−1	1																			
$\{6,1^2\}$	1	−1	−1	1																		
$\{5,3\}$	0	0	−1	0	1																	
$\{5,2,1\}$	0	1	0	−1	−1	1																
$\{5,1^3\}$	−1	1	1	−1	1	−2	1															
$\{4^2\}$	0	0	0	0	−1	0	0	1														
$\{4,3,1\}$	0	0	1	0	0	−1	0	−1	1													
$\{4,2^2\}$	0	0	−1	1	1	−1	0	0	−1	1												
$\{4,2,1^2\}$	0	−1	0	1	0	1	−1	1	−1	−1	1											
$\{4,1^4\}$	1	−1	−1	1	−1	2	−1	−1	2	1	−3	1										
$\{3^2,2\}$	0	0	0	0	−1	1	0	1	−1	−1	0	0	1									
$\{3^2,1^2\}$	0	0	−1	0	1	1	0	0	−1	1	−1	0	−1	1								
$\{3,2^2,1\}$	0	0	1	−1	0	−1	1	−1	3	0	−1	0	−1	−1	1							
$\{3,2,1^3\}$	0	1	0	−1	−1	−1	1	0	0	0	2	−1	2	−1	−2	1						
$\{3,1^5\}$	−1	1	1	−1	2	−2	1	1	−4	−1	3	−1	−2	3	3	−4	1					
$\{2^4\}$	0	0	0	0	−1	2	−1	1	−2	−2	2	0	2	1	−3	0	0	1				
$\{2^3,1^2\}$	0	0	−1	1	1	0	−1	0	−2	2	−1	1	−1	2	1	−2	0	−1	1			
$\{2^2,1^4\}$	0	−1	1	1	0	0	−1	0	2	−2	−1	1	−1	−2	3	2	−1	1	−3	1		
$\{2,1^6\}$	1	−1	−2	1	−2	4	−1	−1	4	3	−6	1	3	−3	−9	8	−1	−1	6	−5	1	
$\{1^8\}$	−1	2	2	−3	2	−6	4	1	−6	−3	12	−5	−3	6	12	−20	6	1	−10	15	−7	1

8. n = 9.

	(9)	$(8,1)$	$(7,2)$	$(7,1^2)$	$(6,3)$	$(6,2,1)$	$(6,1^3)$	$(5,4)$	$(5,3,1)$	$(5,2^2)$	$(5,2,1^2)$	$(5,1^4)$	$(4^2,1)$	$(4,3,2)$	$(4,3,1^2)$	$(4,2^2,1)$	$(4,2,1^3)$	$(4,1^5)$	(3^3)	$(3^2,2,1)$	$(3^2,1^3)$	$(3,2^3)$	$(3,2^2,1^2)$	$(3,2,1^4)$	$(3,1^6)$	$(2^4,1)$	$(2^3,1^3)$	$(2^2,1^5)$	$(2,1^7)$	(1^9)
$\{9\}$	1																													
$\{8,1\}$	−1	1																												
$\{7,2\}$	0	−1	1																											
$\{7,1^2\}$	1	−1	−1	1																										
$\{6,3\}$	0	[illegible]	−1	0	1																									
$\{6,2,1\}$	0	1	0	−1	−1	1																								
$\{6,1^3\}$	−1	1	1	−1	1	−2	1																							
$\{5,4\}$	0	0	0	0	−1	0	0	1																						
$\{5,3,1\}$	0	0	1	0	0	−1	0	−1	1																					
$\{5,2^2\}$	0	0	−1	1	1	−1	0	0	−1	1																				
$\{5,2,1^2\}$	0	−1	0	1	0	1	−1	1	−1	−1	1																			
$\{5,1^4\}$	1	−1	−1	1	−1	2	−1	−1	2	1	−3	1																		
$\{4^2,1\}$	0	0	0	0	1	0	0	−1	−1	0	0	0	1																	
$\{4,3,2\}$	0	0	0	0	−1	1	0	1	0	−1	0	0	−1	1																
$\{4,3,1^2\}$	0	0	−1	0	0	1	0	1	0	1	−1	0	−1	−1	1															
$\{4,2^2,1\}$	0	0	1	−1	0	−1	1	−1	2	0	−1	0	1	−1	−1	1														
$\{4,2,1^3\}$	0	1	0	−1	0	−1	1	−1	−1	0	2	−1	1	2	−1	−2	1													
$\{4,1^5\}$	−1	1	1	−1	1	−2	1	2	−2	−1	3	−1	−2	−2	3	3	−4	1												
$\{3^3\}$	0	0	0	0	0	0	0	0	−1	1	0	0	1	−2	0	0	0	0	1											
$\{3^2,2,1\}$	0	0	0	0	1	−1	0	−1	0	0	1	0	1	1	−1	−1	0	0	−1	1										
$\{3^2,1^3\}$	0	0	1	0	−1	−1	0	0	1	−1	1	0	0	0	−1	2	−1	0	1	−2	1									
$\{3,2^3\}$	0	0	0	0	−1	2	−1	1	−1	−1	1	0	−1	0	2	−1	0	0	1	−2	0	1								
$\{3,2^2,1^2\}$	0	0	−1	1	0	1	−1	1	−1	1	−2	1	−2	0	3	1	−1	0	0	0	−1	−1	1							
$\{3,2,1^4\}$	0	−1	0	1	1	1	−1	0	0	0	−2	1	1	−2	−1	−1	3	−1	−1	4	−1	1	−3	1						
$\{3,1^6\}$	1	−1	−1	1	−2	2	−1	−2	4	1	−3	1	2	4	−6	−3	4	−1	1	−6	4	−1	6	−5	1					
$\{2^4,1\}$	0	0	0	0	1	−2	1	−1	0	1	1	−1	2	0	−4	−1	2	0	−1	4	1	−1	−3	0	0	1				
$\{2^3,1^3\}$	0	0	1	−1	−1	0	1	0	2	−2	1	−1	0	0	−1	2	−2	1	1	−5	2	2	3	−2	0	−2	1			
$\{2^2,1^5\}$	0	1	−1	−1	0	0	1	0	−2	2	1	−1	−1	2	4	−3	−2	1	0	0	−3	−3	3	3	−1	3	−4	1		
$\{2,1^7\}$	−1	1	2	−1	2	−4	1	2	−4	−3	6	−1	−2	−6	6	9	−8	1	−1	9	−4	4	−18	10	−1	−4	10	−6	1	
$\{1^9\}$	1	−2	−2	3	−2	6	−4	−2	6	3	−12	5	3	6	−12	−12	20	−6	1	−12	10	−4	30	−30	7	5	−20	21	−8	1

4. Rules for the analysis of the direct product with illustrative examples. If (ϵ') is the associate partition of n to (ϵ) $\phi_{(\epsilon')}(s)$ is obtained from $\phi_{(\epsilon)}(s)$ by changing the signs of $s_2, s_4, \cdots$ (p. 453). Hence the analysis of $\phi_{(\epsilon')}(s)\phi_{(\nu')}(s)$ follows from that of $\phi_{(\epsilon)}(s)\phi_{(\nu)}(s)$ by taking the associates of all the partitions found in the latter. Thus from

$$\{2,1\}\cdot\{1^2\} = \{3,2\} + \{3,1^2\} + \{2^2,1\} + \{2,1^3\}$$

we read

$$\{2,1\}\cdot\{2\} = \{4,1\} + \{3,2\} + \{3,1^2\} + \{2^2,1\}.$$

In calculating $\{\epsilon\}\cdot\{\nu\}$ it is convenient to take $n \leqq m$ and to have (ϵ) preceded by (if not identical with) its associate (ϵ'); otherwise we calculate first $\{\epsilon'\}\cdot\{\nu'\}$ and read off from this the desired result. The obvious reason for this is that the tables of the preceding paragraph and of (1) pp. 475-477 are simpler as we proceed farther out in our ordered set of partitions. As we have seen in § 2 the general procedure in analysing $\{\epsilon\}\cdot\{\nu\}$ may be formalised as follows

A. Precede by ν_1 each partition of the, supposed known, analysis of $\{\nu_2, \cdots, \nu_m\}$ by $\{\epsilon\}$. This step is the same for all $\{\epsilon\}$, $\{\nu\}$ and we need say nothing further about it.

B. Precede by $\nu_1 + 1$ each partition of the, supposed known, analysis of the product of $\{\nu_2, \cdots, \nu_m\}$ by a linear combination of simple characteristics of the symmetric group on $n - 1$ letters. This linear combination is determined, by the method explained in § 2, from the tables of § **3** and of (1) pp. 475-477. We shall denote it by the symbol B and shall give below tables furnishing B for all partitions (ϵ) of $n \leqq 8$ for which (ϵ) follows, if it is not identical with, its associate (ϵ').

C. Precede by $\nu_1 + 2$ each partition of the, supposed known, analysis of the product of $\{\nu_2, \cdots, \nu_m\}$ by a linear combination of the symmetric group on $n - 2$ letters. We denote this linear combination by C and give it for the same partitions (ϵ) as before.

D. Same as B save that $\nu_1 + 1$ and $n - 1$ are replaced by $\nu_1 + 3$ and $n - 3$, respectively.

E. Same as B save that $\nu_1 + 1$ and $n - 1$ are replaced by $\nu_1 + 4$ and $n - 4$, respectively,

and so on.

Tables furnishing the linear combinations *B*, *C*, *D*, *E*

1. n = 2.

	B
$\{1^2\}$	$\{1\}$

2. n = 3.

	B	C
$\{1^3\}$	$\{1^2\}$	
$\{2,1\}$	$\{2\} + \{1^2\}$	$\{1\}$

3. n = 4.

	B	C
$\{1^4\}$	$\{1^3\}$	
$\{2,1^2\}$	$\{2,1\} + \{1^3\}$	$\{1^2\}$
$\{2^2\}$	$\{2,1\}$	$\{2\}$

4. n = 5.

	B	C	D
$\{1^5\}$	$\{1^4\}$		
$\{2,1^3\}$	$\{2,1^2\} + \{1^4\}$	$\{1^3\}$	
$\{2^2,1\}$	$\{2^2\} + \{2,1^2\}$	$\{2,1\}$	
$\{3,1^2\}$	$\{3,1\} + \{2,1^2\}$	$\{2,1\} + \{1^3\}$	$\{1^2\}$

5. n = 6.

	B	C	D
$\{1^6\}$	$\{1^5\}$		
$\{2,1^4\}$	$\{2,1^3\} + \{1^5\}$	$\{1^4\}$	
$\{2^2,1^2\}$	$\{2^2,1\} + \{2,1^3\}$	$\{2,1^2\}$	
$\{2^3\}$	$\{2^2,1\}$	$\{2^2\}$	
$\{3,1^3\}$	$\{3,1^2\} + \{2,1^3\}$	$\{2,1^2\} + \{1^4\}$	$\{1^3\}$
$\{3,2,1\}$	$\{3,2\} + \{3,1^2\} + \{2^2,1\}$	$\{3,1\} + \{2^2\} + \{2,1^2\}$	$\{2,1\}$

6. n = 7.

	B	C	D	E
$\{1^7\}$	$\{1^6\}$			
$\{2,1^5\}$	$\{2,1^4\} + \{1^6\}$	$\{1^5\}$		
$\{2^2,1^3\}$	$\{2^2,1^2\} + \{2,1^4\}$	$\{2,1^3\}$		
$\{2^3,1\}$	$\{2^3\} + \{2^2,1^2\}$	$\{2^2,1\}$		
$\{3,1^4\}$	$\{3,1^3\} + \{2,1^4\}$	$\{2,1^3\} + \{1^5\}$	$\{1^4\}$	
$\{3,2,1^2\}$	$\{3,2,1\} + \{3,1^3\} + \{2^2,1^2\}$	$\{3,1^2\} + \{2^2,1\} + \{2,1^3\}$	$\{2,1^2\}$	
$\{3,2^2\}$	$\{3,2,1\} + \{2^3\}$	$\{3,2\} + \{2^2,1\}$	$\{2^2\}$	
$\{4,1^3\}$	$\{4,1^2\} + \{3,1^3\}$	$\{3,1^2\} + \{2,1^3\}$	$\{2,1^2\} + \{1^4\}$	$\{1^3\}$

7. n = 8.

	B	C	D	E
$\{1^8\}$	$\{1^7\}$			
$\{2,1^6\}$	$\{2,1^5\} + \{1^7\}$	$\{1^6\}$		
$\{2^2,1^4\}$	$\{2^2,1^3\} + \{2,1^5\}$	$\{2,1^4\}$		
$\{2^3,1^2\}$	$\{2^3,1\} + \{2^2,1^3\}$	$\{2^2,1^2\}$		
$\{2^4\}$	$\{2^3,1\}$	$\{2^3\}$		
$\{3,1^5\}$	$\{3,1^4\} + \{2,1^5\}$	$\{2,1^4\} + \{1^6\}$	$\{1^5\}$	
$\{3,2,1^3\}$	$\{3,2,1^2\} + \{3,1^4\} + \{2^2,1^3\}$	$\{3,1^3\} + \{2^2,1^2\} + \{2,1^4\}$	$\{2,1^3\}$	
$\{3,2^2,1\}$	$\{3,2^2\} + \{3,2,1^2\} + \{2^3,1\}$	$\{3,2,1\} + \{2^3\} + \{2^2,1^2\}$	$\{2^2,1\}$	
$\{3^2,1^2\}$	$\{3^2,1\} + \{3,2,1\}$	$\{3,2,1\} + \{3,1^3\}$	$\{3,1^2\}$	
$\{3^2,2\}$	$\{3^2,1\} + \{3,2^2\}$	$\{3^2\} + \{3,2,1\}$	$\{3,2\}$	
$\{4,1^4\}$	$\{4,1^3\} + \{3,1^4\}$	$\{3,1^3\} + \{2,1^4\}$	$\{2,1^3\} + \{1^5\}$	$\{1^4\}$
$\{4,2,1^2\}$	$\{4,2,1\} + \{4,1^3\} + \{3,2,1^2\}$	$\{4,1^2\} + \{3,2,1\} + \{3,1^3\} + \{2^2,1^2\}$	$\{3,1^2\} + \{2^2,1\} + \{2,1^3\}$	$\{2,1^2\}$

Illustrative examples. Since we give below the complete table analysing the direct product $\Gamma \cdot \Gamma'$ for the cases $m + n = 10$ we shall illustrate by examples for which $m + n > 10$. We arrange the notation so that $n \leq m$ and if (ϵ) precedes its associate partition (ϵ') we first multiply the associated representations and then take the associates of all terms occurring in the product.

Example 1. $\{2, 1\} \cdot \{4, 2^2\}$.

We read from Table 6 (**1**, p. 484)

$$\{2^2\} \cdot \{2, 1\} = \{4, 3\} + \{4, 2, 1\} + \{3^2, 1\} + \{3, 2^2\} + \{3, 2, 1^2\} + \{2^3, 1\}$$

and prefix a 4 to each of the partitions of 7 which occur on the right. From Table 2 above we read $B = \{2\} + \{1^2\}$ and from Table 5 (**1**, p. 484) we read

$$\{2^2\} \cdot [\{2\} + \{1^2\}] = \{4, 2\} + \{3^2\} + 2\{3, 2, 1\} + \{2^3\} + \{2^2, 1^2\};$$

we then prefix a 5 to each of the partitions of 6 which occur on the right. Again from Table 2 above we read $C = \{1\}$ and from Table 4 (**1**, p. 483) $\{2^2\}\{1\} = \{3, 2\} + \{2^2, 1\}$; we prefix a 6 to each of the partitions of 5 which occurs on the right. Hence the desired analysis of the direct product $\Gamma_{(2,1)} \cdot \Gamma_{(4,2^2)}$ is

$$\begin{aligned}\{2, 1\} \cdot \{4, 2^2\} = \{6, 3, 2\} &+ \{6, 2^2, 1\} + \{5, 4, 2\} + \{5, 3^2\} + 2\{5, 3, 2, 1\} \\ &+ \{5, 2^3\} + \{5, 2^2, 1^2\} + \{4^2, 3\} + \{4^2, 2, 1\} + \{4, 3^2, 1\} \\ &+ \{4, 3, 2^2\} + \{4, 3, 2, 1^2\} + \{4, 2^3, 1\}.\end{aligned}$$

As a check against possible errors in copying from the tables the dimensions of the representations on both sides should be calculated. Thus $\{2, 1\}$ is the characteristic of an irreducible representation, of dimension 2, of the symmetric group on 3 letters whilst $\{4, 2^2\}$ is the characteristic of an irreducible representation, of dimension 56, of the symmetric group on 8 letters. Hence, (**1**, p. 448) $\{2, 1\} \cdot \{4, 2^2\}$ is the characteristic of a reducible representation, of dimension $2 \cdot 56 \cdot 11! \div 3!\,8! = 2^4 \cdot 3 \cdot 5 \cdot 7 \cdot 11 = 18{,}480$, of the symmetric group on 11 letters. The dimensions of the various irreducible representations of the symmetric group on 11 letters which occur in the analysis of $\{2, 1\} \cdot \{4, 2^2\}$ given above may be calculated by means of the formula of Frobenius (**1** (11) p. 460) or read off from the table (**7**, pp. 201-204). On removing, for convenience, the common factor 11 we obtain the check

$$1{,}680 = 90 + 100 + 90 + 60 + 420 + 75 \\ + 140 + 42 + 120 + 108 + 120 + 210 + 105.$$

Example 2. $\{3,1\} \cdot \{7,2\}$.

Here $(3,1)$ precedes its associate $(2,1^2)$ and so we calculate $\{2,1^2\} \cdot \{2^2,1^5\}$. Reading the tables as explained in detail in the previous example we have first to calculate $\{2,1^2\} \cdot \{2,1^5\}$. To do this we read off

$$\{2,1^2\}\{1^5\} = \{3,2^2,1^2\} + \{3,2,1^4\} + \{3,1^6\} + \{2^3,1^3\} + \{2^2,1^5\} + \{2,1^7\}.$$

On prefixing a 2 the terms arising from those partitions of 9 on the right which begin with a 3 vanish since $\{2,3,2^2,1^2\} = 0$ (1, p. 461) etc. Hence the terms in the analysis of $\{2,1^2\} \cdot \{2,1^5\}$ which begin with a 2 are $\{2^4,1^3\} + \{2^3,1^5\} + \{2^2,1^7\}$. To find the terms beginning with a 3 we read off

$$[\{2,1\} + \{1^3\}]\{1^5\} = \{3,2,1^3\} + \{3,1^5\} \\ + \{2^3,1^2\} + 2\{2^2,1^4\} + 2\{2,1^6\} + \{1^8\}$$

so that the terms beginning with a 3 are

$$\{3^2,2,1^3\} + \{3^2,1^5\} + \{3,2^3,1^2\} + 2\{3,2^2,1^4\} + 2\{3,2,1^6\} + \{3,1^8\}.$$

The terms beginning with a 4 are found from

$$\{1^2\} \cdot \{1^5\} = \{2^2,1^3\} + \{2,1^5\} + \{1^7\}$$

and are

$$\{4,2^2,1^3\} + \{4,2,1^5\} + \{4,1^7\}.$$

Hence we have

$$\{2,1^2\} \cdot \{2,1^5\} = \{4,2^2,1^3\} + \{4,2,1^5\} + \{4,1^7\} + \{3^2,2,1^3\} + \{3^2,1^5\} \\ + \{3,2^3,1^2\} + 2\{3,2^2,1^4\} + 2\{3,2,1^6\} + \{3,1^8\} \\ + \{2^4,1^3\} + \{2^3,1^5\} + \{2^2,1^7\}$$

(a preliminary calculation which would have been unnecessary if we had already prepared the table for $n + m = 11$). The terms beginning with a 2 in the analysis of $\{2,1^2\} \cdot \{2^2,1^5\}$ are accordingly

$$\{2,4,2^2,1^3\} + \{2,4,2,1^5\} + \{2,4,1^7\} + \{2^5,1^3\} + \{2^4,1^5\} + \{2^3,1^7\} \\ = -\{3^2,2^2,1^3\} - \{3^2,2,1^5\} - \{3^2,1^7\} + \{2^5,1^3\} + \{2^4,1^5\} + \{2^3,1^7\}$$

(the terms beginning with 3, in the analysis of $\{2, 1^2\} \cdot \{2, 1^5\}$ vanishing when the 2 is prefixed). To find the terms beginning with 3 we read off, from the table below, $[\{2, 1\} + \{1^3\}] \cdot \{2, 1^5\}$ obtaining, on prefixing a 3,

$$\begin{aligned}\{3^3, 1^4\} + 2\{3^2, 2^2, 1^3\} + 3\{3^2, 2, 1^5\} + 2\{3^2, 1^7\}& \\ + \{3, 2^4, 1^2\} + 2\{3, 2^3, 1^4\} + 2\{3, 2^2, 1^6\} + \{3, 2, 1^8\}&.\end{aligned}$$

To find the terms beginning with 4 we read off $\{1^2\} \cdot \{2, 1^5\}$ and obtain, on prefixing a 4,

$$\{4, 3, 2, 1^4\} + \{4, 2^3, 1^3\} + \{4, 3, 1^6\} + \{4, 2^2, 1^5\} + \{4, 2, 1^7\}.$$

Collecting our results we find

$$\begin{aligned}\{2, 1^2\} \cdot \{2^2, 1^5\} = {}& \{4, 3, 2, 1^4\} + \{4, 2^3, 1^3\} + \{4, 3, 1^6\} + \{4, 2^2, 1^5\} \\ & + \{4, 2, 1^7\} + \{3^3, 1^4\} + \{3^2, 2^2, 1^3\} + 2\{3^2, 2, 1^5\} \\ & + \{3^2, 1^7\} + \{3, 2^4, 1^2\} + 2\{3, 2^3, 1^4\} + 2\{3, 2^2, 1^6\} \\ & + \{3, 2, 1^8\} + \{2^5, 1^3\} + \{2^4, 1^5\} + \{2^3, 1^7\}.\end{aligned}$$

On taking the associates of the representations occurring in this analysis we obtain

$$\begin{aligned}\{3, 1\} \cdot \{7, 2\} = {}& \{10, 3\} + \{10, 2, 1\} + \{9, 4\} + 2\{9, 3, 1\} + \{9, 2^2\} \\ & + \{9, 2, 1^2\} + \{8, 5\} + 2\{8, 4, 1\} + 2\{8, 3, 2\} + \{8, 3, 1^2\} \\ & + \{8, 2^2, 1\} + \{7, 5, 1\} + \{7, 4, 2\} + \{7, 4, 1^2\} \\ & + \{7, 3^2\} + \{7, 3, 2, 1\}.\end{aligned}$$

On dividing out by the common factor 13 we have the check by dimensions:

$$\begin{aligned}4455 = {}& 16 + 33 + 33 + 210 + 72 + 110 + 44 + 396 \\ & + 528 + 324 + 280 + 220 + 462 + 528 + 275 + 924.\end{aligned}$$

Example 3. $\{2^2, 1\} \cdot \{4, 3, 1\}$.

This is the example, referred to in the introduction, of which the analysis was wrongly printed in (2). We read off

$$\begin{aligned}\{2^2, 1\} \cdot \{3, 1\} = {}& \{5, 3, 1\} + \{5, 2^2\} + \{5, 2, 1^2\} + \{4, 3, 2\} + \{4, 3, 1^2\} \\ & + 2\{4, 2^2, 1\} + \{4, 2, 1^3\} + \{3^2, 2, 1\} + \{3, 2^3\} + \{3, 2^2, 1^2\}\end{aligned}$$

so that the terms in the desired analysis beginning with 4 are

$$\begin{aligned}\{4^2, 3, 2\} + \{4^2, 3, 1^2\} + 2\{4^2, 2^2, 1\} + \{4^2, 2, 1^3\}& \\ + \{4, 3^2, 2, 1\} + \{4, 3, 2^3\} + \{4, 3, 2^2, 1^2\}&.\end{aligned}$$

Also we read off

$$\begin{aligned}[\{2^2\} + \{2, 1^2\}] \cdot \{3, 1\} = {} & \{5, 3\} + 2\{5, 2, 1\} + \{5, 1^3\} + 2\{4, 3, 1\} \\ & + 2\{4, 2^2\} + 3\{4, 2, 1^2\} + \{4, 1^4\} + \{3^2, 2\} \\ & + \{3^2, 1^2\} + 2\{3, 2^2, 1\} + \{3, 2, 1^3\}\end{aligned}$$

and the terms in the desired analysis beginning with 5 are obtained by prefixing a 5 to the partitions of 8 which appear on the right-hand side. Finally

$$\begin{aligned}\{2, 1\} \cdot \{3, 1\} = {} & \{5, 2\} + \{5, 1^2\} + \{4, 3\} + 2\{4, 2, 1\} + \{4, 1^3\} \\ & + \{3^2, 1\} + \{3, 2^2\} + \{3, 2, 1^2\}\end{aligned}$$

and the terms in the desired analysis beginning with 6 are found by prefixing a 6 to the partitions of 7 which appear on the right-hand side. Collecting terms we obtain

$$\begin{aligned}\{2^2, 1\} \cdot \{4, 3, 1\} = {} & \{6, 5, 2\} + \{6, 5, 1^2\} + \{6, 4, 3\} + 2\{6, 4, 2, 1\} \\ & + \{6, 4, 1^3\} + \{6, 3^2, 1\} + \{6, 3, 2^2\} + \{6, 3, 2, 1^2\} \\ & + \{5^2, 3\} + 2\{5^2, 2, 1\} + \{5^2, 1^3\} + 2\{5, 4, 3, 1\} \\ & + 2\{5, 4, 2^2\} + 3\{5, 4, 2, 1^2\} + \{5, 4, 1^4\} + \{5, 3^2, 2\} \\ & + \{5, 3^2, 1^2\} + 2\{5, 3, 2^2, 1\} + \{5, 3, 2, 1^3\} + \{4^2, 3, 2\} \\ & + \{4^2, 3, 1^2\} + 2\{4^2, 2^2, 1\} + \{4^2, 2, 1^3\} + \{4, 3^2, 2, 1\} \\ & + \{4, 3, 2^3\} + \{4, 3, 2^2, 1^2\}.\end{aligned}$$

On dividing out by the common factor 13×11 we obtain the check by dimensions

$$\begin{aligned}3150 = {} & 36 + 40 + 45 + 240 + 72 + 80 + 84 + 144 + 24 + 120 + 35 \\ & + 210 + 180 + 450 + 63 + 81 + 112 + 300 + 144 + 60 + 81 \\ & + 180 + 84 + 105 + 60 + 120.\end{aligned}$$

Example 4. $\{2^4\} \cdot \{4^2\}$.

For the first set of terms we have to evaluate $\{2^4\} \cdot \{4\}$ and for this we calculate its associate:

$$\{1^4\} \cdot \{4^2\} = \{5^2, 1^2\} + \{5, 4, 1^3\} + \{4^2, 1^4\}$$

so that

$$\{2^4\} \cdot \{4\} = \{6, 2^3\} + \{5, 2^3, 1\} + \{4, 2^4\}.$$

The first set of terms follows by prefixing a 4 and is — $\{5^2, 2^3\} + \{4^2, 2^4\}$. The second set of terms requires the evaluation of $\{2^3, 1\} \cdot \{4\}$; the associate of this,

$$\{1^4\} \cdot \{4, 3\} = \{5, 4, 1^2\} + \{5, 3, 1^3\} + \{4^2, 1^3\} + \{4, 3, 1^4\}$$

so that

$$\{2^3, 1\} \cdot \{4\} = \{6, 2^2, 1\} + \{5, 2^3\} + \{5, 2^2, 1^2\} + \{4, 2^3, 1\}.$$

Hence the second group of terms is

$$\{5^2, 2^3\} + \{5^2, 2^2, 1^2\} + \{5, 4, 2^3, 1\}.$$

The third and last set of terms requires the product

$$\{2^3\} \cdot \{4\} = \{6, 2^2\} + \{5, 2^2, 1\} + \{4, 2^3\}$$

and is

$$\{6^2, 2^2\} + \{6, 5, 2^2, 1\} + \{6, 4, 2^3\}.$$

Collecting we have

$$\begin{aligned}\{2^4\} \cdot \{4^2\} = \{6^2, 2^2\} &+ \{6, 5, 2^2, 1\} + \{6, 4, 2^3\} \\ &+ \{5^2, 2^2, 1^2\} + \{5, 4, 2^3, 1\} + \{4^2, 2^4\}.\end{aligned}$$

A partial check on the accuracy of the analysis is furnished by the fact that the six representations occurring on the right consist of three pairs of associated representations so that the direct product $\Gamma_{(2^4)} \cdot \Gamma_{(4^2)}$ is self-associated, as it must be. On removing the common factor $13 \cdot 11 \cdot 7 \cdot 5 \cdot 3 \cdot 2$ we obtain the check by dimensions $84 = 5 + 21 + 16 + 16 + 21 + 5$. The direct product being analysed in this example is a representation, of dimension $13 \cdot 11 \cdot 7^2 \cdot 5 \cdot 3^2 \cdot 2^3 = 2{,}522{,}520$ of the symmetric group on 16 letters.

5. Table furnishing the analysis of $\Gamma . \Gamma'$ for $n + m = 10$. (For an explanation of how this table is read see **1**, p. 482).

	(10)	$(9,1)$	$(8,2)$	$(8,1^2)$	$(7,3)$	$(7,2,1)$	$(7,1^3)$	$(6,4)$	$(6,3,1)$	$(6,2^2)$	$(6,2,1^2)$	$(6,1^4)$	(5^2)	$(5,4,1)$	$(5,3,2)$	$(5,3,1^2)$	$(5,2^2,1)$	$(5,2,1^3)$	$(5,1^5)$	$(4^2,2)$	$(4^2,1^2)$	$(4,3^2)$	$(4,3,2,1)$	$(4,3,1^3)$	$(3^3,1)$	
$(9).(1)$	1	1																								$(1^9).(1)$
$(8,1).(1)$		1	1	1																						$(2,1^7).(1)$
$(7,2).(1)$			1		1	1																				$(2^2,1^5).(1)$
$(7,1^2).(1)$				1		1	1																			$(3,1^6).(1)$
$(6,3).(1)$					1			1	1																	$(2^3,1^3).(1)$
$(6,2,1).(1)$						1			1	1	1															$(3,2,1^4).(1)$
$(6,1^3).(1)$							1				1	1														$(4,1^5).(1)$
$(5,4).(1)$								1					1	1												$(2^4,1).(1)$
$(5,3,1).(1)$									1				1	1	1											$(3,2^2,1^2).(1)$
$(5,2^2).(1)$										1					1		1									$(3^2,1^3).(1)$
$(5,2,1^2).(1)$											1					1	1	1								$(4,2,1^3).(1)$
$(5,1^4).(1)$												1						1	1							$(5,1^4).(1)$
$(4^2,1).(1)$														1						1	1					$(3,2^3).(1)$
$(4,3,2).(1)$															1					1		1	1			$(3^2,2,1).(1)$
$(4,3,1^2).(1)$																1					1		1	1		$(4,2^2,1).(1)$
$(3^3).(1)$																						1			1	$(3^3).(1)$
$(8).(2)$	1	1	1																							$(1^8).(1^2)$
$(7,1).(2)$		1	1	1	1	1																				$(2,1^6).(1^2)$
$(6,2).(2)$			1		1	1		1	1	1																$(2^2,1^4).(1^2)$
$(6,1^2).(2)$				1		1	1		1		1															$(3,1^5).(1^2)$
$(5,3).(2)$					1			1	1				1	1	1											$(2^3,1^2).(1^2)$
$(5,2,1).(2)$						1			1	1	1			1	1	1	1									$(3,2,1^3).(1^2)$
$(5,1^3).(2)$							1				1	1				1		1								$(4,1^4).(1^2)$
$(4^2).(2)$								1						1						1						$(2^4).(1^2)$
$(4,3,1).(2)$									1					1	1	1				1	1	1	1			$(3,2^2,1).(1^2)$
	(1^{10})	$(2,1^8)$	$(2^2,1^6)$	$(3,1^7)$	$(2^3,1^4)$	$(3,2,1^5)$	$(4,1^6)$	$(2^4,1^2)$	$(3,2^2,1^3)$	$(3^2,1^4)$	$(4,2,1^4)$	$(5,1^5)$	(2^5)	$(3,2^3,1)$	$(3^2,2,1^2)$	$(4,2^2,1^2)$	$(4,3,1^3)$	$(5,2,1^3)$	$(6,1^4)$	$(3^2,2^2)$	$(4,2^3)$	$(3^3,1)$	$(4,3,2,1)$	$(5,2^2,1)$	$(4,3^2)$	

	(10)	$(9,1)$	$(8,2)$	$(8,1^2)$	$(7,3)$	$(7,2,1)$	$(7,1^3)$	$(6,4)$	$(6,3,1)$	$(6,2^2)$	$(6,2,1^2)$	$(6,1^4)$	(5^2)	$(5,4,1)$	$(5,3,2)$	$(5,3,1^2)$	$(5,2^2,1)$	$(5,2,1^3)$	$(5,1^5)$	$(4^2,2)$	
$(4,2^2).(2)$										1					1		1			1	
$(4,2,1^2).(2)$											1					1	1	1			
$(4,1^4).(2)$												1						1	1		
$(3^2,2).(2)$															1						
$(3^2,1^2).(2)$																1					
$(3,2^2,1).(2)$																	1				
$(3,2,1^3).(2)$																		1			
$(3,1^5).(2)$																			1		
$(2^4).(2)$																					
$(2^3,1^2).(2)$																					
$(2^2,1^4).(2)$																					
$(2,1^6).(2)$																					
$(1^8).(2)$																					
$(7).(3)$	1	1	1		1																
$(6,1).(3)$		1	1	1	1	1		1	1												
$(5,2).(3)$			1		1	1		1	1	1			1	1	1						
$(5,1^2).(3)$				1		1	1		1		1			1		1					
$(4,3).(3)$					1			1	1					1	1					1	
$(4,2,1).(3)$						1			1	1	1			1	1	1	1			1	1
$(4,1^3).(3)$							1				1	1				1		1			1
$(3^2,1).(3)$									1						1	1					
$(3,2^2).(3)$										1					1		1				
$(3,2,1^2).(3)$											1					1	1	1			
$(3,1^4).(3)$												1						1	1		
$(2^3,1).(3)$																	1				
$(2^2,1^3).(3)$																		1			
$(2,1^5).(3)$																			1		
$(1^7).(3)$																					
$(7).(2,1)$		1	1	1		1															
$(6,1).(2,1)$			1	1	1	2	1		1	1	1										
$(5,2).(2,1)$					1	1		1	2	1	1			1	1	1	1				
$(5,1^2).(2,1)$						1	1		1	1	2	1			1	1	1	1			
$(4,3).(2,1)$								1	1				1	2	1	1				1	1
$(4,2,1).(2,1)$									1	1	1			1	2	2	2	1		1	1
$(4,1^3).(2,1)$											1	1				1	1	2	1		
$(3^2,1).(2,1)$														1	1	1				1	1
$(6).(4)$	1	1	1		1			1													
$(5,1).(4)$		1	1	1	1	1		1	1				1	1							
$(4,2).(4)$			1		1	1		1	1	1				1	1					1	
$(4,1^2).(4)$				1		1	1		1		1			1		1					1
$(3^2).(4)$					1				1						1						
$(3,2,1).(4)$						1			1	1	1				1	1	1				
$(3,1^3).(4)$							1				1	1				1		1			
$(2^3).(4)$										1							1				
$(2^2,1^2).(4)$											1						1	1			
$(2,1^4).(4)$												1						1	1		
$(1^6).(4)$																			1		
$(6).(3,1)$		1	1	1	1	1			1												
$(5,1).(3,1)$			1	1	1	2	1	1	2	1	1			1	1	1					
$(4,2).(3,1)$					1	1		1	2	1	1		1	2	2	1	1			1	1
$(4,1^2).(3,1)$						1	1		1	1	2	1		1	1	2	1	1		1	1
$(3^2).(3,1)$								1	1					1	1	1				1	
$(3,2,1).(3,1)$									1	1	1			1	2	2	2	1		1	1
$(3,1^3).(3,1)$											1	1				1	1	2	1		1
$(2^3).(3,1)$															1		1				
	(1^{10})	$(2,1^8)$	$(2^2,1^6)$	$(3,1^7)$	$(2^3,1^4)$	$(3,2,1^5)$	$(4,1^6)$	$(2^4,1^2)$	$(3,2^2,1^3)$	$(3^2,1^4)$	$(4,2,1^4)$	$(5,1^5)$	(2^5)	$(3,2^3,1)$	$(3^2,2,1^2)$	$(4,2^2,1^2)$	$(4,3,1^3)$	$(5,2,1^3)$	$(6,1^4)$	$(3^2,2^2)$	

	$(4,3^2)$	$(4,3,2,1)$	$(4,3,1^3)$	$(4,2^3)$	$(4,2^2,1^2)$	$(4,2,1^4)$	$(4,1^6)$	$(3^3,1)$	$(3^2,2^2)$	$(3^2,2,1^2)$	$(3^2,1^4)$	$(3,2^3,1)$	$(3,2^2,1^3)$	$(3,2,1^5)$	$(3,1^7)$	(2^5)	$(2^4,1^2)$	$(2^3,1^4)$	$(2^2,1^6)$	$(2,1^8)$	(1^{10})
$(3^2,1^2).(1^2)$		1		1																	
$(4,2,1^2).(1^2)$		1	1		1																
$(5,1^3).(1^2)$			1			1															
$(3^2,2).(1^2)$	1	1						1	1												
$(4,2^2).(1^2)$		1	1					1		1											
$(4,3,1).(1^2)$		1		1	1				1	1		1									
$(5,2,1).(1^2)$			1		1	1				1	1		1								
$(6,1^2).(1^2)$						1	1				1			1							
$(4^2).(1^2)$				1								1				1					
$(5,3).(1^2)$					1							1	1				1				
$(6,2).(1^2)$						1							1	1				1			
$(7,1).(1^2)$							1							1	1				1		
$(8).(1^2)$															1					1	
$(1^7).(1^3)$																					
$(2,1^5).(1^3)$																					
$(2^2,1^3).(1^3)$																					
$(3,1^4).(1^3)$																					
$(2^3,1).(1^3)$	1																				
$(3,2,1^2).(1^3)$		1																			
$(4,1^3).(1^3)$			1																		
$(3,2^2).(1^3)$	1	1						1													
$(3^2,1).(1^3)$		1		1					1												
$(4,2,1).(1^3)$		1	1		1					1											
$(5,1^2).(1^3)$			1			1					1										
$(4,3).(1^3)$				1	1							1									
$(5,2).(1^3)$					1	1							1								
$(6,1).(1^3)$						1	1							1							
$(7).(1^3)$							1								1						
$(1^7).(2,1)$																					
$(2,1^5).(2,1)$																					
$(2^2,1^3).(2,1)$																					
$(3,1^4).(2,1)$																					
$(2^3,1).(2,1)$		1																			
$(3,2,1^2).(2,1)$	1	2	1	1	1																
$(4,1^3).(2,1)$		1	1		1	1															
$(3,2^2).(2,1)$	1	2	1					1	1	1											
$(1^6).(1^4)$																					
$(2,1^4).(1^4)$																					
$(2^2,1^2).(1^4)$																					
$(3,1^3).(1^4)$																					
$(2^3).(1^4)$	1																				
$(3,2,1).(1^4)$		1																			
$(4,1^2).(1^4)$			1																		
$(3^2).(1^4)$				1																	
$(4,2).(1^4)$					1																
$(5,1).(1^4)$						1															
$(6).(1^4)$							1														
$(1^6).(2,1^2)$																					
$(2,1^4).(2,1^2)$																					
$(2^2,1^2).(2,1^2)$	1	1																			
$(3,1^3).(2,1^2)$		1	1																		
$(2^3).(2,1^2)$	1	1						1													
$(3,2,1).(2,1^2)$	1	3	1	1	1			1	1	1											
$(4,1^2).(2,1^2)$		1	2		1	1				1	1										
$(3^2).(2,1^2)$		1		1	1				1			1									
	$(3^3,1)$	$(4,3,2,1)$	$(5,2^2,1)$	$(4^2,1^2)$	$(5,3,1^2)$	$(6,2,1^2)$	$(7,1^3)$	$(4,3^2)$	$(4^2,2)$	$(5,3,2)$	$(6,2^2)$	$(5,4,1)$	$(6,3,1)$	$(7,2,1)$	$(8,1^2)$	(5^2)	$(6,4)$	$(7,3)$	$(8,2)$	$(9,1)$	(10)

	(10)	$(9,1)$	$(8,2)$	$(8,1^2)$	$(7,3)$	$(7,2,1)$	$(7,1^3)$	$(6,4)$	$(6,3,1)$	$(6,2^2)$	$(6,2,1^2)$	$(6,1^4)$	(5^2)	$(5,4,1)$	$(5,3,2)$	$(5,3,1^2)$	$5,2^2,1)$	$(5,2,1^3)$	$(5,1^5)$	$(4^2,2)$	$(4^2,1^2)$
$(2^2,1^2).(3,1)$																1	1	1			
$(2,1^4).(3,1)$																		1	1		
$(1^6).(3,1)$																					
$(6).(2^2)$			1			1				1											
$(5,1).(2^2)$					1	1			1	1	1				1		1				
$(4,2).(2^2)$								1	1	1				1	1	1	1			1	
$(4,1^2).(2^2)$									1		1				1	1	1	1			
$(3^2).(2^2)$													1	1	1						1
$(3,2,1).(2^2)$														1	1	1	1			1	1
$(5).(5)$	1	1	1		1			1					1								
$(4,1).(5)$		1	1	1	1	1		1	1					1							
$(3,2).(5)$			1		1	1			1	1					1						
$(3,1^2).(5)$				1		1	1		1		1					1					
$(2^2,1).(5)$						1				1	1						1				
$(2,1^3).(5)$							1				1	1						1			
$(1^5).(5)$												1							1		
$(4,1).(4,1)$			1	1	1	2	1	1	2	1	1		1	2	1	1				1	1
$(3,2).(4,1)$					1	1		1	2	1	1			1	2	1	1			1	
$(3,1^2).(4,1)$						1	1		1	1	2	1		1	1	2	1	1			1
$(2^2,1).(4,1)$									1	1	1				1	1	2	1			
$(2,1^3).(4,1)$											1	1				1	1	2	1		
$(3,2).(3,2)$								1	1	1			1	2	2	1	1			1	1
$(3,1^2).(3,2)$									1		1			1	1	2	1	1		1	1
$(2^2,1).(3.2)$														1	1	1	1			1	1
$(3,1^2).(3,1^2)$										1	1	1			1	1	2	2	1	1	1
	(1^{10})	$(2,1^8)$	$(2^2,1^6)$	$(3,1^7)$	$(2^3,1^4)$	$(3,2,1^5)$	$(4,1^6)$	$(2^4,1^2)$	$(3,2^2,1^3)$	$(3^2,1^4)$	$(4,2,1^4)$	$(5,1^5)$	(2^5)	$(3,2^3,1)$	$(3^2,2,1^2)$	$(4,2^2,1^2)$	$(4,3,1^3)$	$(5,2,1^3)$	$(6,1^4)$	$(3^2,2^2)$	$(4,2^3)$

REFERENCES.

1. F. D. Murnaghan, "On the representations of the symmetric group," *American Journal of Mathematics*, **59** (1937), pp. 437-488.
2. D. E. Littlewood and A. R. Richardson, "Group characters and algebra," *Philosophical Transactions of the Royal Society of London* (A), **233** (1934), pp. 99-141.
3. C. Kostka, "Über den Zusammenhang zwischen einigen Formen von symmetrischen Funktionen," *Journal für die reine und angewandte Mathematik* (Crelle), **93** (1882), pp. 89-123.
4. ———, "Tafeln und Formeln für symmetrische Funktionen," *Jahrbuch der Deutschen math. Ver.*, **16** (1907), pp. 429-451.
5. ———, "Tafeln für symmetrische Funktionen bis zur elften Dimension mit kurzen Erläuterungen," *Prog.* (5) *kgl. Gym. u. Realgymn. Insterburg* (1908).
6. I. Schur, "Über die rationalen Darstellungen der allgemeinen linearen Gruppe," *Berliner Berichte* (1927), pp. 58-75.
7. M. Zia-ud-Din, "The characters of the symmetric group of order 11!," *Proceedings of the London Mathematical Society* (2), **39** (1935), pp. 200-204.
8. J. R. Roe, "Interfunctional expressibility tables of symmetric functions," Syracuse University (1931). This, privately published, collection of tables contains (Plate 1) Kostka's table for $n = 10$.

$(4,3^2)$	$(4,3,2,1)$	$(4,3,1^3)$	$(4,2^3)$	$(4,2^2,1^2)$	$(4,2,1^4)$	$(4,1^6)$	$(3^3,1)$	$(3^2,2^2)$	$(3^2,2,1^2)$	$(3^2,1^4)$	$(3,2^3,1)$	$(3,2^2,1^3)$	$(3,2,1^5)$	$(3,1^7)$	(2^5)	$(2^4,1^2)$	$(2^3,1^4)$	$(2^2,1^6)$	$(2,1^8)$	(1^{10})	
	1	1	1	2	1				1		1	1									$(4,2).(2,1^2)$
		1		1	2	1				1		1	1								$(5,1).(2,1^2)$
					1	1							1	1							$(6).(2,1^2)$
																					$(1^6).(2^2)$
																					$(2,1^4).(2^2)$
	1		1																		$(2^2,1^2).(2^2)$
1	1			1																	$(3,1^3).(2^2)$
	1							1													$(2^3).(2^2)$
1	2	1	1	1			1	1	1		1										$(3,2,1).(2^2)$
																					$(1^5).(1^5)$
																					$(2,1^3).(1^5)$
																					$(2^2,1).(1^5)$
																					$(3,1^2).(1^5)$
																					$(3,2).(1^5)$
																					$(4,1).(1^5)$
																					$(5).(1^5)$
																					$(2,1^3).(2,1^3)$
1	1																				$(2^2,1).(2,1^3)$
	1	1																			$(3,1^2).(2,1^3)$
	1		1	1																	$(3,2).(2,1^3)$
		1		1	1																$(4,1).(2,1^3)$
1	2		1				1	1													$(2^2,1).(2^2,1)$
1	2	1		1			1		1												$(3,1^2).(2^2,1)$
	2	1	1	1				1	1		1										$(3,2).(2^2,1)$
	2	2	1	1	1			1	1	1											$(3,1^2).(3,1^2)$
$(3^3,1)$	$(4,3,2,1)$	$(5,2^2,1)$	$(4^2,1^2)$	$(5,3,1^2)$	$(6,2,1^2)$	$(7,1^3)$	$(4,3^2)$	$(4^2,2)$	$(5,3,2)$	$(6,2^2)$	$(5,4,1)$	$(6,3,1)$	$(7,2,1)$	$(8,1^2)$	(5^2)	$(6,4)$	$(7,3)$	$(8,2)$	$(9,1)$	(10)	

ON A CLASS OF ARITHMETICAL FOURIER SERIES.*

By PHILIP HARTMAN.

Let

$$(1)\qquad \psi(x) = -\frac{1}{\pi}\sum_{k=1}^{\infty}\frac{\sin 2k\pi x}{k} = \begin{cases} x-[x]-\frac{1}{2} & \text{if } x \neq [x] \\ 0 & \text{if } x = [x] \end{cases},$$

then we have the formal identity

$$(2)\qquad \sum_{n=1}^{\infty}\frac{c_n\psi(nx)}{n} = -\frac{1}{\pi}\sum_{k=1}^{\infty}\frac{1}{k}\left(\sum_{d/k} c_d\right)\sin 2k\pi x.$$

There have been several papers justifying this process for particular sequences $\{c_n\}$; see, for example, Landau [1] for $c_n = n^{-1}$; Chowla [2] for $c_n = n^{\alpha}$, $\alpha < 0$; Chowla and Walfisz [3] for $c_n = 1$; Hartman and Wintner [4] for $c_n = n^{\alpha}$, $\alpha < \frac{1}{2}$; Davenport [5] for $c_n = \mu(n)$, for $c_n = \lambda(n)$, for $c_n = \Lambda(n)$, and, finally, for $c_{n^2} = \mu(n)$ and all other $c_n = 0$. In most of the above examples, (2) is shown to hold almost everywhere, but (2) is valid for all x in the cases treated by Landau,[1] Chowla,[2] and the last example of Davenport.[5]

A more general problem is treated in this paper. Let $f(x)$ be a function of period 1 which can be represented almost everywhere by a convergent trigonometrical series, say

$$(3)\qquad f(x) = \tfrac{1}{2}a_0 + \sum_{k=1}(a_k\cos 2k\pi x + b_k \sin 2k\pi x).$$

Conditions on $f(x)$ and on sequences $\{c_n\}$ will be investigated under which the following identity

$$(4)\qquad \sum_{n=1}^{\infty} c_n f(nx) = \tfrac{1}{2}a_0\sum_{k=1}^{\infty} c_k + \sum_{k=1}^{\infty}\left[\left(\sum_{d/k} c_d a_{k/d}\right)\cos 2k\pi x + \left(\sum_{d/k} c_d b_{k/d}\right)\sin 2k\pi x\right]$$

---

* Received July 28, 1937.

[1] E. Landau, "Konvergenzbeweis einer Lerchschen Reihe," *Mémoires de la Société Royale des Sciences de Bohême, Classe des Sciences*, (1919), IV.

[2] S. D. Chowla, "Some problems in diophantine approximation (I)," *Mathematische Zeitschrift*, vol. 33 (1931), pp. 544-563.

[3] S. D. Chowla and A. Walfisz, "Über eine Riemannsche Identität," *Acta Arithmetica*, vol. 1 (1935), pp. 87-112.

[4] P. Hartman and A. Wintner, "On certain Fourier series involving sums of divisors," to appear in the *Acta Arithmetica*.

[5] H. Davenport, "On some infinite series involving arithmetical functions," *Quarterly Journal of Mathematics*, vol. 8 (1937), pp. 8-13.

is valid. Landau, Chowla, Walfisz, and Davenport did not seem to recognize that the problem was a Fourier series problem and obtained their results with the aid of diophantine approximation. Fourier series methods are employed here; the methods are essentially the same as those used by Wintner.[6]

THEOREM 1. *Let $f(x)$ be a Lebesgue integrable function with the period* 1 *and*

$$f(x) \sim \sum_{k=1}^{\infty} (a_k \cos 2k\pi x + b_k \sin 2k\pi x). \tag{5}$$

Let α and β be two real numbers such that

$$m = \min (\alpha, \beta) > \tfrac{1}{2}, \tag{6}$$

and

$$a_k = O(k^{-\alpha}), \qquad b_k = O(k^{-\alpha}), \tag{7}$$

and, finally,

$$c_n = O(n^{-\beta}). \tag{8}$$

Then

(i) *the partial sums of the series*

$$\sum_{n=1}^{\infty} c_n f(nx) \tag{9}$$

tend in the mean (L^q) to a function $F(x)$, for every $q < 1/(1-m)$ or for every $q > 0$ according as $m < 1$ or $m \geqq 1$;

(ii) *$F(x)$ possesses the Fourier series*

$$F(x) \sim \sum_{k=1}^{\infty} \left[\left(\sum_{d/k} c_d a_{k/d} \right) \cos 2k\pi x + \left(\sum_{d/k} c_d b_{k/d} \right) \sin 2k\pi x \right]; \tag{10}$$

(iii) *the Fourier series of $F(x)$ converges for almost all x to $F(x)$,*

$$F(x) = \sum_{k=1}^{\infty} \left[\left(\sum_{d/k} c_d a_{k/d} \right) \cos 2k\pi x + \left(\sum_{d/k} c_d b_{k/d} \right) \sin 2k\pi x \right]. \tag{11}$$

It follows from the formal identity (4) that it was necessary to suppose $a_0 = 0$, unless $\sum_{n=1}^{\infty} c_n$ is a convergent series. The conditions imposed on $f(x)$ allow several immediate conclusions to be drawn. Suppose, first, that $\alpha < 1$ and let p be a number such that

$$2 > p > 1/\alpha,$$

---

[6] A. Wintner, "On a trigonometrical series of Riemann," *American Journal of Mathematics*, vol. 59 (1937), pp. 629-634.

so that

$$\sum_{k=1}^{\infty} (|a_k|^p + |b_k|^p) < \infty;$$

then by the Hausdorff[7] extension of the Fischer-Riesz theorem, $f(x)$ belongs to the class $L^{p'}$, $p' = p/(p-1) < 1/(1-\alpha)$. By the same arguments, it follows that if $\alpha \geqq 1$, $f(x)$ belongs to L^q for all $q > 0$. Thus, in any case, $f(x)$ belongs to L^2. Also there exists a positive number δ such that

$$2\alpha - \delta > 1,$$

hence

$$\sum_{k=1}^{\infty} (a_k^2 + b_k^2) k^{\delta} < \infty,$$

this implies that the Fourier series (5) is convergent almost everywhere to $f(x)$,

$$f(x) = \sum_{k=1}^{\infty} (a_k \cos 2k\pi x + b_k \sin 2k\pi x). \tag{12}$$

Proof of (*i*). Suppose first that $m < 1$, it is to be proved that if

$$F_N(x) = \sum_{n=1}^{N} c_n f(nx), \tag{13}$$

then there exists a function $F(x)$ such that

$$\int_0^1 [F(x) - F_N(x)]^q dx \to 0 \text{ as } N \to \infty \text{ if } q < 1/(1-m). \tag{14}$$

Now it is clear that $F_N(x)$ is integrable and

$$F_N(x) \sim \sum_{k=1}^{\infty} \left[\left(\sum_{d/k}^{d \leqq N} c_d a_{k/d} \right) \cos 2k\pi x + \left(\sum_{d/k}^{d \leqq N} c_d b_{k/d} \right) \sin 2k\pi x \right]. \tag{15}$$

Let $M > N$, then it follows from the Hausdorff[7] extension of the Parseval relation that

$$\int_0^1 [F_M(x) - F_N(x)]^q dx \leqq \left[\sum_{k=1}^{\infty} \left| \sum_{d/k}^{N < d \leqq M} c_d a_{k/d} \right|^{q/(q-1)} + \sum_{k=1}^{\infty} \left| \sum_{d/k}^{N < d \leqq M} c_d b_{k/d} \right|^{q/(q-1)} \right]^{q-1} \tag{16}$$

provided that $q \geqq 2$, as one may suppose. Thus, if $q \geqq 2$,

---

[7] F. Hausdorff, "Eine Ausdehnung des Parsevalschen Satzes über Fourierreihen," *Mathematische Zeitschrift*, vol. 16 (1923), pp. 163-169.

(17) $$\int_0^1 [F_M(x) - F_N(x)]^q dx \leqq \left[\sum_{k=N}^{\infty} \left(\sum_{d/k} | c_d a_{k/d} | \right)^{q/(q-1)} + \sum_{k=N}^{\infty} \left(\sum_{d/k} | c_d b_{k/d} | \right)^{q/(q-1)} \right]^{q-1}.$$

Now from (7) and (8)

(18) $$\sum_{d/k} | c_d a_{k/d} | = O\left(\frac{1}{k^{\alpha}} \sum_{d/k} d^{\alpha-\beta} \right) = O\left(\frac{\sigma_{\alpha-\beta}(k)}{k^{\alpha}} \right),$$

where

(19) $$\sigma_{\gamma}(k) = \sum_{d/k} d^{\gamma}.$$

It is known[8] that if $\gamma \geqq 0$, then

(20) $$\sigma_{\gamma}(k) = O(k^{\gamma+\epsilon}),$$

and if $\gamma \leqq 0$, then

(21) $$\sigma_{\gamma}(k) = O(k^{\epsilon}),$$

where $\epsilon > 0$ is arbitrary. Thus from (6), (18), (20) and (21), it follows that

(22) $$\sum_{d/k} | c_d a_{k/d} | = O(k^{-m+\epsilon}),$$

and similarly

(23) $$\sum_{d/k} | c_d b_{k/d} | = O(k^{-m+\epsilon}).$$

Hence if $q < 1/(1-m)$, then $q/(q-1) > 1/m$, so that the series

(24) $$\sum_{k=1}^{\infty} \left[\left(\sum_{d/k} | c_d a_{k/d} | \right)^{q/(q-1)} + \left(\sum_{d/k} | c_d b_{k/d} | \right)^{q/(q-1)} \right]$$

is convergent. Thus, from (17),

(25) $$\int_0^1 [F_M(x) - F_N(x)]^q dx \to 0, \text{ as } M, N \to \infty$$

if $q < 1/(1-m)$. Thus, for the case $m < 1$, (i) now follows from (25) and the completeness of the space L^q. The case where $m \geqq 1$ follows similarly.

Proof of (ii). This follows at once from (i), for if a_k^N, b_k^N are the k-th Fourier coefficients of $F_N(x)$, then it follows from (14) that

(26) $$\lim_{N\to\infty} a_k^N = A_k \quad \text{and} \quad \lim_{N\to\infty} b_k^N = B_k$$

exist and are the k-th Fourier coefficients of $F(x)$. Now from (15),

(27) $$a_k^N = \sum_{d/k}^{d \leqq N} c_d a_{k/d} \quad \text{and} \quad b_k^N = \sum_{d/k}^{d \leqq N} c_d b_{k/d},$$

---

[8] T. H. Gronwall, "Some asymptotic expressions in the theory of numbers," *Transactions of the American Mathematical Society*, vol. 14 (1913), pp. 113-122.

so that (10) follows from (26) and (27).

Proof of (iii). From (10), (22), and (23), it follows that

$$A_k = O(k^{-m+\epsilon}) \quad \text{and} \quad B_k = O(k^{-m+\epsilon}), \tag{28}$$

so that there exists a $\delta > 0$ such that

$$\sum_{k=1}^{\infty} (A_k^2 + B_k^2) k^\delta < \infty, \tag{29}$$

for $\epsilon > 0$ and $\delta > 0$ can be chosen so that

$$2m - 2\epsilon - \delta > 1.$$

That the equality (11) is valid almost everywhere is a consequence of (10) and (29).

Next, there will be proved[9]

THEOREM 2. *If $f(x)$ satisfies the conditions of Theorem 1 when the inequality (6) is replaced by*

$$m > \tfrac{2}{3} \tag{30}$$

and if $f(x)$ is bounded, then the series (9) converges almost everywhere to $F(x)$.

Proof. Let p be a number satisfying

$$0 < p < 3, \tag{31}$$

then, in virtue of (30),

$$p(1 - m) < 1. \tag{32}$$

Also there exists an $\epsilon > 0$ and a number p satisfying (31), and a fortiori (32), such that

$$p(2m - 1 - \epsilon) > 1. \tag{33}$$

---

[9] This theorem and a similar proof for the case where $f(x)$ is the function (1) and where $c_n = n^{-a}$, $a > 2/3$ was also known to Professor Walfisz (unpublished), as I understood after this paper was completed. Incidentally, a similar theorem was proved by F. Jerosch and H. Weyl, "Über die Konvergenz von Reihen, die nach periodischen Funktionen fortschreiten," *Mathematische Annalen*, vol. 66 (1909), pp. 67-80.

(Cf., also S. Chowla, "On some infinite series involving arithmetical functions" I, II, *Proceedings of the Indian Academy of Sciences*, Section A, vol. 5 (1937), pp. 511-516. These references were obtained from the *Zentralblatt für Mathematik*, vol. 17 (1937), p. 5 as the above journal was not available. Added Nov. 18, 1937.)

In view of (10) and (15)

$$\int_0^1 [F(x)-F_N(x)]^2dx \leqq \sum_{k=1}^{\infty}\left[\left(\sum_{\substack{d>N\\ d/k}} c_d a_{k/d}\right)^2+\left(\sum_{\substack{d>N\\ d/k}} c_d b_{k/d}\right)^2\right],$$

so that from (22) and (23),

$$(34)\qquad \int_0^1 [F(x)-F_N(x)]^2dx = O\Big(\sum_{k=N+1}^{\infty} k^{-2m+\epsilon}\Big) = O(N^{-2m+1+\epsilon}).$$

It is seen that if p satisfies (33), then

$$\sum_{N=1}^{\infty}\int_0^1 [F(x)-F_{[N^p]}(x)]^2dx$$

is a convergent series; thus, it follows from Fatou's lemma that

$$(35)\qquad \sum_{N=1}^{\infty}[F(x)-F_{[N^p]}(x)]^2$$

is convergent almost everywhere. Hence for almost all x

$$(36)\qquad F_{[N^p]}(x)\to F(x) \text{ as } N\to\infty \text{ if } p \text{ satisfies (33).}$$

Now let j be an integer such that

$$(37)\qquad [N^p]\leqq j<[(N+1)^p],$$

then

$$F_j(x)-F_{[N^p]}(x)=\sum_{n=[N^p]}^{j} c_n f(nx) = O\Big(\sum_{n=[N^p]}^{[(N+1)^p]} n^{-m}\Big),$$

or

$$(38)\qquad F_j(x)-F_{[N^p]}(x) = O(N^{-pm}\cdot N^{p-1}) = O(N^{p(1-m)-1}).$$

Hence, if j satisfies (37),

$$(39)\qquad F_j(x)-F_{[N^p]}(x)\to 0 \text{ as } N\to\infty \text{ if } p \text{ satisfies (32).}$$

Thus (36), (39) and the existence of a p satisfying (32) and (33) implies the statement of Theorem 2.

From this theorem, the following result can be inferred: Let $f(x)$ be a bounded integrable function of period 1 such that

$$f(x)\sim a_0/2+\sum_{k=1}^{\infty}(a_k\cos 2k\pi x+b_k\sin 2k\pi x)$$

and

$$a_k=O(k^{-(2/3)-\epsilon}),\qquad b_k=O(k^{-(2/3)-\epsilon}),$$

then for almost all x

$$S_N(x;f)=\sum_{n=1}^{N} f(nx)-N\int_0^1 f(t)\,dt = O(N^{2/3+\epsilon}).$$

The theorems just proved imply all of the earlier results mentioned at the beginning of the paper except those which replaced " almost all x " by " all x." To obtain such results, it is clear that the conditions on the sequence $\{c_n\}$ must be such as to insure the convergence of the series (9) everywhere, thus defining $F(x)$ by an everywhere convergent series instead of using a definition which leaves the function undetermined on a zero set, as in Theorem 1 or 2. It is also clear that $f(x)$ will have to satisfy more stringent conditions than those above. In this connection, one has

THEOREM 3. *Let the series*

$$\sum_{n=1}^{\infty} c_n \tag{40}$$

be absolutely convergent. Let $f(x)$ be a bounded integrable function of period 1 such that

$$g(u,x) = \frac{1}{u}\int_0^u [(f(x+t) + f(x-t) - 2f(x)]dt, \qquad u > 0, \tag{41}$$

is of bounded variation in any u-interval to the right of $u = 0$ (when considered as a function of u) for all x and such that

$$g(u,x) \to 0, \qquad u \to 0 \tag{42}$$

for all x. Let $V_n(t,x)$ denote the total variation of $g(nu, nx)$ in the interval $0 < u < t$ and suppose that

$$\sum_{n=1}^{\infty} | c_n | V_n(t,x) \tag{43}$$

is convergent for some $t = t(x) > 0$. The identity

$$\sum_{n=1}^{\infty} c_n f(nx) = \tfrac{1}{2} a_0 \sum_{n=1}^{\infty} c_n + \sum_{k=1}^{\infty} \left[\left(\sum_{d/k} c_d a_{k/d} \right) \cos 2k\pi x + \left(\sum_{d/k} c_d b_{k/d} \right) \sin 2k\pi x \right] \tag{44}$$

is then valid for all x.

Proof. This theorem is merely a rewriting of the de la Vallée Poussin test for the convergence of Fourier series for the case at hand. In fact, the conditions that $g(u,x)$ be of bounded variation in some interval $0 < u < a$ and that $g(u,x) \to 0$ as $u \to 0$ are precisely the conditions required by the de la Vallée Poussin test for the convergence of the Fourier series of $f(x)$ at the point x. Thus the conditions of the theorem imply that (3) is valid for all x.

That

$$(45)\quad F(x) = \sum_{n=1}^{\infty} c_n f(nx)$$

$$\sim \tfrac{1}{2} a_0 \sum_{k=1}^{\infty} c_n + \sum_{n=1}^{\infty} \left[\left(\sum_{d/k} c_d a_{k/d} \right) \cos 2k\pi x + \left(\sum_{d/k} c_d b_{k/d} \right) \sin 2k\pi x \right]$$

is an immediate consequence of the uniform convergence of the series (9), which, in turn, follows from the boundedness of $f(x)$ and the absolute convergence of the series (40).

One obtains from (41) that

$$g(nu, nx) = \frac{1}{nu} \int_0^{nu} [f(nx + t) + f(nx - t) - 2f(nx)] dt.$$

If the integration variable t is changed to nt, one has

$$(46)\quad g(nu, nx) = \frac{1}{u} \int_0^{u} [f(nx + nt) + f(nx - nt) - 2f(nx)] dt.$$

Now put

$$(47)\quad G(u, x) = \frac{1}{u} \int_0^{u} [F(x + t) + F(x - t) - 2F(x)] dt,$$

then it is clear from (46) that

$$(48)\quad G(u, x) = \sum_{n=1}^{\infty} c_n g(nu, nx).$$

If $t = t(x) > 0$ is chosen so that the series (43) is convergent, $G(u, x)$ is of bounded variation in the interval $0 < u < t$, since the variation of the sum of functions does not exceed the sum of the variations of the functions. Also since $g(u, x)$ is a bounded function, the series in (48) is uniformly convergent, thus it follows from

$$g(nu, nx) \rightarrow 0, \qquad u \rightarrow 0,$$

that

$$(49)\quad G(u, x) \rightarrow 0, \qquad u \rightarrow 0.$$

Thus the function $F(x)$ satisfies the criterion of de la Vallée Poussin at every point, hence (44) is valid at every point in virtue of (45). This completes the proof.

Theorem 3 implies the very particular cases treated by Landau, Chowla, and Davenport. Let the function $f(x)$ of Theorem 3 be the function (1) and let $g(u, x)$ have the corresponding meaning. Let n and x be fixed; suppose that x is not of the form $j/2n$, where j is any integer, otherwise $g(nu, nx) = 0$ for all u. Let

$$\frac{k-1}{n} < x < \frac{k}{n},$$

and put

$$\alpha_1 = \max\left(x - \frac{k-1}{n}, \frac{k}{n} - x\right), \qquad \alpha_2 = \frac{1}{n} - \alpha_1. \tag{50}$$

Then

$$\psi(nx+nt) + \psi(nx-nt) - 2\psi(nx) = 0$$
$$\text{if } 0 < t < \alpha_2;$$

$$\psi(nx+nt) + \psi(nx-nt) - 2\psi(nx) = (-1)^{\mu(x)} \tag{51}$$
$$\text{if } r/n + \alpha_2 < t < r/n + \alpha_1; \qquad (r = 0, 1, \cdots),$$
$$\psi(nx+nt) + \psi(nx-nt) - 2\psi(nx) = 0$$
$$\text{if } r/n + \alpha_1 < t < (r+1)/n + \alpha_2;$$

where $\mu(x)$ is 0 or 1 according as $x < (2k-1)/2n$ or $x > (2k-1)/2n$. Hence

$$g(nu, nx) = 0$$
$$\text{if } 0 < u < \alpha_2$$

$$g(nu, nx) = (-1)^{\mu(x)}[r(\alpha_1 - \alpha_2) + u - r/n - \alpha_2]/u \tag{52}$$
$$\text{if } r/n + \alpha_2 < u < r/n + \alpha_1; \qquad (r = 0, 1, \cdots).$$
$$g(nu, nx) = (-1)^{\mu(x)}(r+1)(\alpha_1 - \alpha_2)/u$$
$$\text{if } r/n + \alpha_1 < u < (r+1)/n + \alpha_2;$$

It follows from these relations, by straightforward appraisals which can be left to the reader, that

$$V_n(t, x) \leqq C \log nt, \quad \text{if} \quad t > \frac{1}{n} \tag{53}$$

where C is a constant independent of n, t and x. Also, it is obvious that (42) is satisfied. Thus, if the sequence $\{c_n\}$ is such that

$$\sum_{n=1}^{\infty} \left| \frac{c_n \log n}{n} \right| \tag{54}$$

is convergent, the identity (2) is valid for *all* x.

THE JOHNS HOPKINS UNIVERSITY.

THE STRUCTURE OF LOCAL CLASS FIELD THEORY.*

By O. F. G. Schilling.†

The gist of local class field theory may be sought in the characterization of the finite abelian extensions over perfect fields whose valuation groups are discrete and archimedian and whose residue class fields are finite Galois fields. M. Moriya and the author of the present paper recently investigated the generalization of this theory to infinite extensions of $\mathfrak{p}$-adic number fields.[1] We found that the structural theory of finite abelian extensions over such infinite fields is in general determined by the nature of the value group and the field of residue classes. The results obtained can be interpreted from a more complex viewpoint as a theory of sufficient conditions which admit the development of a local class field theory securing the same theorems as in the finite case. Thus one is naturally lead to the following problem which we wish to investigate in this paper: *Is it possible to prove structural properties of a field which is perfect with respect to a discrete archimedian valuation if some set of standard theorems of local class field theory is valid for the given ground field?* Or, we ask for necessary conditions of theorems in local class field theory.

We shall see that most of the standard theorems imply that the residue field of the given perfect field is algebraically perfect[2] and that it possesses for each integer n exactly one cyclic extension of degree n. One easily observes that the set of possible residue class fields thus described is rather extensive, thus it is necessary to impose further conditions upon the given ground field if one wishes to characterize the $\mathfrak{p}$-adic number fields we investigated in the paper already mentioned. The simplest assumption we found was to postulate the validity of the given set of theorems for all perfect subfields of the given field.

In order to avoid unnecessary repetitions in our later investigations it is advisable to recall some known facts about the algebraic and arithmetical

---

* Received September 28, 1937.

† Johnston Scholar at The Johns Hopkins University for 1937-38.

[1] M. Moriya and O. F. G. Schilling, "Zur Klassenkörpertheorie über unendlichen perfekten Körpern," *Journal of the Fac. of Sci. Hokk. Imp. Univ.*, ser. I, vol. 5 (Sapporo, Japan) 1937.

[2] We shall use "algebraically perfect" as translation of the German term "vollkommen."

theory of perfect fields. Let k be a field which is perfect with respect to a discrete archimedian valuation $\mathfrak{p}$. The maximal order of all $\mathfrak{p}$-adic integers in k be denoted by $\mathfrak{o}$, and let $\mathfrak{p} = (\pi)$ be the prime ideal of $\mathfrak{o}$. Then the ring $\mathfrak{o}/\mathfrak{p}$ of residue classes is a field $\boldsymbol{k}$. In case that $\boldsymbol{k}$ is an algebraically perfect field the structure of k is uniquely determined by that of $\boldsymbol{k}$ provided that the characteristic $\chi(\boldsymbol{k})$ is a unit or prime element of k.[3] There can be established a one to one correspondence between the finite algebraic extensions $\boldsymbol{U}$ of $\boldsymbol{k}$ and the unramified extensions U of k, the fields $\boldsymbol{U}$ being the residue class fields of the fields U. In particular, if U is a normal extension with the Galois group $\Gamma(U, k)$ then the Galois group $\Gamma(\boldsymbol{U}, \boldsymbol{k})$ of the residue class field is equal to $\Gamma(U, k)$.

The arithmetic theory of normal division algebras D of finite rank m over k is very simple because of the validity of Hensel's criterion of reducibility.[4] All quantities of D which satisfy a minimal equation with highest coefficient one and whose coefficients lie in $\mathfrak{o}$ form a maximal order $\mathfrak{O}$, these elements are characterized by the property that their reduced norms are $\mathfrak{p}$-adic integers. The ring $\mathfrak{O}$ contains a uniquely determined principal prime ideal $\mathfrak{P}$ whose powers exhaust the set of all ideals with respect to $\mathfrak{O}$. The ring of residue classes $\mathfrak{O}/\mathfrak{P}$ is a division algebra of finite rank f over the field $\mathfrak{o}/\mathfrak{p}$, f being called the residue degree of D with respect to k.[5] Since $\mathfrak{O}$ is a principal ring the extension $\mathfrak{O}\mathfrak{p} = \mathfrak{p}\mathfrak{O}$ of the prime ideal $\mathfrak{p}$ is a positive power $\mathfrak{P}^e$ of the two-sided prime ideal $\mathfrak{P}$; the integer e is called the ramification degree of D with respect to k. Between the two numbers e and f related to D with respect to k the following important relation holds

$$ef = m.$$

Since a prime element Π of $\mathfrak{P}$ generates a ramified subfield $k(\Pi)$ of D whose ramification exponent is equal to e we find that e is a divisor of the degree $n = m^{1/2}$ of D.[6] Hence

$$e \leqq n \quad \text{and} \quad f \geqq n.$$

A normal division algebra D of degree n over k can always be represented in its class of normal algebras over k as the crossed product $(a(\sigma,\tau), U, \Gamma(U,k))$ belonging to a suitably chosen normal unramified splitting field U of D. For an arbitrary but fixed choice of a prime element π of $\mathfrak{p}$ this crossed product decomposes uniquely into a cyclic ramified algebra and an unramified algebra

---

[3] H. Hasse and F. K. Schmidt, "Die Struktur diskret bewerteter Körper," *Crelle*, vol. 170 (1933).

[4] H. Hasse, "Über $\mathfrak{p}$-adische Schiefkörper und ihre Bedeutung für die Arithmetik hyperkomplexer Zahlsysteme," *Mathematische Annalen*, vol. 104 (1931).

[5] M. Deuring, "Algebren," *Ergebnisse der Mathematik und ihrer Grenzgebiete*, Berlin 1935, Chap. VI, §§ 11, 12.

[6] T. Nakayama, "Divisionalgebren über diskret bewerteten perfekten Körpern," *Crelle*, vol. 177, 1937.

$$(*) \qquad D \sim (a(\sigma,\tau), U, \Gamma(U,k)) \sim (\pi^{e(\sigma,\tau)}, U, \Gamma(U,k)) \times (\eta(\sigma,\tau), U, \Gamma(U,k)),$$

where $a(\sigma,\tau) = \pi^{e(\sigma,\tau)}\eta(\sigma,\tau)\epsilon(\sigma,\tau)$; $\eta(\sigma,\tau)$ belonging to a fixed set of multiplicative representatives of the residue class field $\boldsymbol{U}$ and $\epsilon(\sigma,\tau) \equiv 1 \pmod{\mathfrak{p}}$. Since the exponents $e(\sigma,\tau)$ form a set of addends belonging to the Galois group $\Gamma(U,k)$ they uniquely determine a cyclic subfield of U. It is readily seen that the algebra $(\pi^{e(\sigma,\tau)}, U, \Gamma(U,k))$ is similar to a ramified division algebra possessing the aforementioned cyclic subfield of U as splitting field. The second factor in (*) represents an unramified division algebra D' which corresponds to a division algebra $\boldsymbol{D}'$ over $\boldsymbol{k}$; we observe that $\boldsymbol{D}'$ need not be normal over $\boldsymbol{k}$.[7]

A simple group theoretical consideration shows that the group of all ramified algebras $\{(\pi^{e(\sigma,\tau)}, U, \Gamma(U,k))\}$ split by a fixed normal unramified field U over k is isomorphic with the factor group $\Gamma(U,k)/\Gamma(U,k)'$ of $\Gamma(U,k)$ with respect to its commutator group.

A normal division algebra D possesses always unramified maximally commutative subfields U provided that the algebra of residue classes $\mathfrak{D}/\mathfrak{P}$ has a center which is separable over the field $\boldsymbol{k}$.[8]

We begin our investigations with the proof of some lemmas which will be quite useful later on.

LEMMA 1. *If the ground field k and its finite algebraic extensions K possess no cyclic unramified extensions then there do not exist proper division algebras of finite rank over k,* i. e. *k is quasi-algebraically closed.*[9]

Proof. Suppose that D is a proper normal division algebra of rank $m = n^2$ over k. The algebra D possesses always separable normal splitting fields U:

$$D \times U \sim U.$$

---

[7] E. Witt, "Schiefkörper über diskret bewerteten Körpern," *Crelle*, vol. 176 (1937).

[8] Cf. the paper mentioned under [6]. In the following investigation we shall assume that all division algebras considered possess residue algebras with separable centers over $\boldsymbol{k}$.

[9] A field is called quasi-algebraically closed if there exist no proper division algebras over it.

Take such a splitting field and consider a Sylow subgroup Σ_q belonging to a prime divisor $q \neq 1$ of n. Let the subfield of U belonging to it be U_q, $([U_q : k], q) = 1$. Since Σ_q is a solvable group we can draw a composition chain

$$1 = \Sigma_q^{(s)} < \Sigma_q^{(s-1)} < \cdots < \Sigma_q^{(i)} < \Sigma_q^{(i-1)} < \cdots < \Sigma_q^{(2)} < \Sigma_q^{(1)} = \Sigma_q.$$

whose factors are cyclic groups of order q. The corresponding chain of fields may be given by

$$\begin{aligned} U = U_q^{(s)} > U_q^{(s-1)} > U_q^{(s-2)} > \cdots \\ > U_q^{(i)} > U_q^{(i-1)} > \cdots > U_q^{(2)} > U_q^{(1)} = U_q. \end{aligned}$$

Now consider the algebra $D \times U_q^{(s-1)}$, it possesses the cyclic unramified field $U = U_q^{(s)}$ as splitting field. Since such a field cannot exist according to our assumptions we see that $D \times U_q^{(s-1)}$ must be similar to $U_q^{(s-1)}$. Repeating this conclusion we observe that already

$$D \times U_q \sim U_q.$$

This last relation contradicts the choice of q for a well known theorem in the algebraic theory of normal algebras asserts that the degree of a splitting field of D must be a multiple of n, and we have shown that D—if it would exist—possesses U_q as splitting field.

LEMMA 2. *If k and all its finite algebraic extensions are never centers of cyclic proper algebras then k is quasi-algebraically closed.*

Proof. We can apply again the same argument, here the relation $D \times U_q \sim U_q$ is a consequence of the supposed non-existence of cyclic algebras over U_q.

LEMMA 3. *If all possibly existing division algebras $\boldsymbol{D}$ of degree n over $\boldsymbol{k}$ and its finite algebraic extensions $\boldsymbol{K}$ possess isomorphic maximally commutative subfields and if -1 is a universal quadratic norm in case that $\chi(\boldsymbol{k}) \neq 2$ then $\boldsymbol{k}$ is quasi-algebraically closed.*[10]

Proof. Suppose that there exists a proper normal division algebra $\boldsymbol{D}$ of degree $n = \Pi q_i^{\sigma_i}$ over $\boldsymbol{k}$. The algebra $\boldsymbol{D}$ is similar to a direct product of primary division algebras $\boldsymbol{D}_i$ whose degrees are equal to $q_i^{\sigma_i}$. In order to prove that our assumptions imply $\boldsymbol{D} \sim \boldsymbol{k}$ it suffices to show that all primary algebras $\boldsymbol{D}_i$ are similar to $\boldsymbol{k}$. First we consider division algebras whose degrees are relatively prime to 2.

---

[10] An element is said to be a universal quadratic norm if it is the norm of a suitable element in any quadratic extension of the ground field.

Let $\boldsymbol{D} = (\boldsymbol{a}, \boldsymbol{Z}_r)$ be a cyclic division algebra of degree r. Then $\boldsymbol{a}^r$ is the least power of $\boldsymbol{a}$ which is norm of an element in $\boldsymbol{Z}_r$. Hence $\boldsymbol{k}(\boldsymbol{a}^{1/r})$ is an algebraic extension of $\boldsymbol{k}$ of degree r, since $\boldsymbol{a}$ is a factor set the field $\boldsymbol{k}(\boldsymbol{a}^{1/r})$ is a splitting field of $\boldsymbol{D}$. Our assumption yields that $\boldsymbol{Z}_r$ and $\boldsymbol{k}(\boldsymbol{a}^{1/r})$ are equal if taken in one and the same algebraically closed field over $\boldsymbol{k}$. But now the element $\boldsymbol{a}$ is the norm of a quantity in $\boldsymbol{Z}_r$ for r was supposed to be relatively prime to 2. Consequently $\boldsymbol{D}$ has to be similar to $\boldsymbol{k}$. Next we apply Lemma 2 which asserts that there do not exist proper division algebras over $\boldsymbol{k}$ as center of these primary algebras. Hence all primary algebras whose degrees are relatively prime to 2 are similar to $\boldsymbol{k}$.

If the degree of a primary algebra is a power of 2 we have to distinguish two different cases:

$$\text{(i) } \chi(\boldsymbol{k}) = 2 \quad \text{and} \quad \text{(ii) } \chi(\boldsymbol{k}) \neq 2.$$

Case (i). According to the theory of primary division algebras and their subalgebras it is sufficient to prove that there do not exist proper division algebras of quaternions over $\boldsymbol{k}$ and its finite algebraic extensions for our assumption is to hold over $\boldsymbol{k}$ and its extensions, and the argument using a Sylow subgroup belonging to the prime 2 of the Galois group of a normal splitting field applies here too.

Suppose now that there exists a proper division algebra of degree 2 over $\boldsymbol{k}$ as center. Such an algebra $\boldsymbol{Q}$ has obviously the form

$$\boldsymbol{Q} = (\boldsymbol{a}, \boldsymbol{k}(\boldsymbol{C}))$$

where $\boldsymbol{C}^2 - \boldsymbol{C} = \boldsymbol{b} \neq 0$, $\boldsymbol{u}^2 = \boldsymbol{a}$, $\boldsymbol{u}^{-1}\boldsymbol{C}\boldsymbol{u} = \boldsymbol{C} + 1$ with $\boldsymbol{a}, \boldsymbol{b} \neq 0$ in $\boldsymbol{k}$. Then our assumption implies that $\boldsymbol{k}(\boldsymbol{u}) = \boldsymbol{k}(\boldsymbol{C})$ in one and the same algebraic closure of $\boldsymbol{k}$. Such an equality is impossible for $\boldsymbol{k}(\boldsymbol{u})$ is an inseparable extension of $\boldsymbol{k}$ and $\boldsymbol{k}(\boldsymbol{C})$ is a separable extension.

Case (ii). It suffices again to show that there do not exist proper division algebras of quaternions over $\boldsymbol{k}$ and its finite algebraic extensions. Let $\boldsymbol{Q} = (\boldsymbol{a}, \boldsymbol{k}(\boldsymbol{b}^{1/2}))$ be such an algebra. Our assumptions imply first of all that $\boldsymbol{k}(\boldsymbol{a}^{1/2}) = \boldsymbol{k}(\boldsymbol{b}^{1/2})$, i. e. $\boldsymbol{b} = \boldsymbol{a}\boldsymbol{c}^2$ with $\boldsymbol{c}$ in $\boldsymbol{k}$. Hence $\boldsymbol{Q} = (\boldsymbol{a}, \boldsymbol{k}(\boldsymbol{a}^{1/2}))$. The norm of $\boldsymbol{a}^{1/2}$ is equal to $-\boldsymbol{a}$. But $\boldsymbol{a}$ itself is also a norm since our special assumption, $-\mathbf{1} = N\boldsymbol{d}$ with $\boldsymbol{d}$ in $\boldsymbol{k}(\boldsymbol{a}^{1/2})$, implies $\boldsymbol{a} = (-1)(-\boldsymbol{a}) = N(\boldsymbol{d}\boldsymbol{a}^{1/2})$.

Remark. If we omit in the assumptions that $-\mathbf{1}$ be a universal quadratic norm then we see that there exists exactly one division algebra over $\boldsymbol{k}$, namely the algebra $(-1, \boldsymbol{k}(-1^{1/2}))$. For the existence of a division algebra $(\boldsymbol{a}, \boldsymbol{k}(\boldsymbol{a}^{1/2}))$ implies that $\boldsymbol{a}$ is not the norm of an element in $\boldsymbol{k}(\boldsymbol{a}^{1/2})$ hence $-\mathbf{1}$ is not a norm and a fortiori not a square. The algebra $(-\mathbf{1}, \boldsymbol{k}(\boldsymbol{a}^{1/2}))$ must be equivalent to $(\boldsymbol{a}, \boldsymbol{k}(\boldsymbol{a}^{1/2}))$ as the assumption on the isomorphism of the maximally commutative subfields implies, consequently

$$(\boldsymbol{a}, \boldsymbol{k}(\boldsymbol{a}^{1/2})) = (-\mathbf{1}, \boldsymbol{k}(-\mathbf{1}^{1/2})).$$

THEOREM 1. *If the field of residue classes $\boldsymbol{k}$ of the perfect field k is algebraically perfect the three following statements are equivalent:*

(i) *the ramification degree e of each normal division algebra D of degree n over k equals n,*

(ii) *the algebra of residue classes belonging to each normal division algebra D over k is a commutative field, and*

(iii) *the maximally commutative ununified subfields of each division algebra D over k are isomorphic and, in case of $\chi(\boldsymbol{k}) \neq 2$, -1 is a universal quadratic norm in $\boldsymbol{k}$.*

Proof. If $e = f = n$ for a division algebra D then

$$\begin{aligned} D &\sim (\pi^{e(\sigma,\tau)}, U, \Gamma(U, k))(\eta(\sigma, \tau), U, \Gamma(U, k)) \\ &\sim (\pi^{e(\sigma,\tau)}, U, \Gamma(U, k)) \sim (\pi, Z_n)^{\nu}, \end{aligned}$$

where Z_n is a maximally commutative cyclic unramified subfield of D and ν is relatively prime to n. Moreover, the algebra has exponent n. The maximal order $\mathfrak{O}$ of D contains the maximal order $\mathfrak{O}(Z_n)$ of Z_n and

$$\mathfrak{O}/\mathfrak{P} \supseteq \mathfrak{O}(Z_n)/\mathfrak{P} \cong \mathfrak{O}(Z_n)/\mathfrak{p}.$$

Since $[\mathfrak{O}/\mathfrak{P} : \boldsymbol{k}] = n$ we have

$$\mathfrak{O}/\mathfrak{P} = \mathfrak{O}(Z_n)/\mathfrak{P},$$

i. e. the algebra of residue classes is a commutative field.

Conversely, let D be a division algebra over k which satisfies (ii). Since $\boldsymbol{k}$ was assumed to be algebraically perfect the algebra D can be represented as a generalized crossed product

$$D = (a(\alpha, \beta, \gamma), U)$$

where U is a suitable maximally commutative unramified subfield of D and α, β, γ are elements of the Galois group belonging to the normal field containing U. Now (ii) implies

$$\mathfrak{O}/\mathfrak{P} \supseteq \mathfrak{O}(U)/\mathfrak{P} \cong \mathfrak{O}(U)/\mathfrak{p}.$$

If the algebra $\mathfrak{O}/\mathfrak{P}$ has the degree n' over its center $\boldsymbol{K}$ and $\boldsymbol{K}$ has the rank n'' over $\boldsymbol{k}$ the following relation holds between the different ranks

$$f = [\mathfrak{O}/\mathfrak{P} : \boldsymbol{k}] = n'^2 n'' = n'' = n.$$

Hence (i) and (ii) are equivalent statements.

Obviously (i) and (ii) imply that all maximally commutative unramified subfields U_n of division algebras D of degree n are isomorphic. Hence $\boldsymbol{k}$ is quasi-algebraically closed and -1 is a universal quadratic norm in $\boldsymbol{k}$, thus (iii) holds. Conversely, (iii) implies according to Lemma 3 that $\boldsymbol{k}$ is quasi-algebraically closed. But this fact implies $e = f = n$. Thus (i) and (iii) are equivalent.

Now let k be a perfect field for which any one of the statements of Theorem 1 holds. We wish to show by an example that such an assumption does not necessarily imply that the well known theorems of the local class field theory hold.[11]

Let $\boldsymbol{k}$ be the field of all roots of unity over the field of all rational numbers. Then the field $k = \boldsymbol{k}\{t\}$ consisting of all formal power series $\sum_{i>-\infty}^{\infty} \boldsymbol{a}_i t^i$ in one variable t with coefficients $\boldsymbol{a}_i$ in $\boldsymbol{k}$ is a perfect field. The unramified extensions U of k are fields of formal power series in t whose coefficients are taken from a field which is isomorphic with the residue field $\boldsymbol{U}$ of U. Let U be a normal unramified field with the Galois group $\Gamma(U, k) = \{\sigma, \tau, \cdots\}$. Then the group $G(U)$ consisting of all normal algebras over k which are split by U is isomorphic with the factor group $\Gamma(U, k)/\Gamma(U, k)'$ of the Galois group $\Gamma(U, k)$ with respect to its commutator group for each algebra in $G(U)$ has the form

$$(\pi^{e(\sigma,\tau)}, U, \Gamma(U, k)),$$

where π denotes a fixed prime element of p. This is true because $\boldsymbol{k}$ is a quasi-algebraically closed field.[12] Moreover all division algebras D over k are cyclic.

Now let U be in particular a normal unramified extension of k whose Galois group is not solvable, such fields always exist. Namely, we have only to take a normal extension $\boldsymbol{U}'$ over the field of all rational numbers and to form the join $\boldsymbol{U} \neq \boldsymbol{U'k}$.[18] The group of algebras belonging to such a field U is equal to one.

---

[11] For general statements about perfect fields also in later considerations see W. Krull, "Allgemeine Bewertungstheorie," *Crelle*, vol. 167 (1931).

[12] H. Hasse, "Die Struktur der R. Brauerschen Algebrenklassengruppe über einem algebraischen Zahlkörper," *Mathematische Annalen*, vol. 107 (1933).

If U is an abelian extension then $G(U)$ is isomorphic with the Galois group $\Gamma(U,k)$. In particular, the group $G(Z_n)$ belonging to a cyclic unramified extension Z_n consists of a single cycle $Z(n)$ of order n.

Consider now the set of all division algebras D of degree and exponent n, these algebras can all be represented as cyclic crossed products

$$(\pi, Z_n)^\nu,$$

where the exponents ν are relatively prime to n. Observe that all algebras (a, Z_n) whose factor sets are units in k, are similar to k. Thus the cyclic ramified extension $k(\pi^{1/n})$ of degree n over k is the universal splitting field of all algebras of degree n

$$G(k(\pi^{1/n})) \supseteq \text{all } G(Z_n).$$

A simple argument using π-adic approximations yields that $G(k(\pi^{1/n}))$ is an infinite abelian group of type $(n, n, \cdots)$.

Furthermore, we see that the factor group of norm classes k^*/NU^* belonging to an unramified abelian extension of degree n is a cyclic group of order n whose generator is representable by t. In order to prove this assertion one has only to represent U as the join of a set of mutually distinct cyclic subfields and to determine the respective factor groups in a composition chain of U.

Combining these results we observe

(i) the group of classes of algebras possessing degree and exponent n is infinite,

(ii) the group of algebras $G(U)$ split by an unramified field is a cycle of order n if and only if U is a cyclic field of degree n,

(iii) the factor group of norm classes associated with an abelian unramified extension of degree n is a cycle of order n.

Consequently most of the standard theorems of local class field theory do not hold for fields k supposing only that the residue fields $\boldsymbol{k}$ are quasi-algebraically closed. Thus we see that it is necessary to impose further conditions upon k in order that the old theory can be re-established.

THEOREM 2. *If the residue class field* $\boldsymbol{k}$ *of the perfect field* k *is quasi-*

---

[18] Such fields are for example the fields whose group is an alternating group of more than 4 variables.

algebraically closed and if for each unramified normal extension U_n *of* k *there exists at least one normal division algebra* D_n *of degree* n *over* k *which is split by* U_n *then all unramified extensions of* k *are cyclic.*

Proof. The first assumption implies that the group of algebras split by U_n is isomorphic with the factor group $\Gamma(U_n, k)/\Gamma(U_n, k)'$. The other assumption $D_n \times U_n \sim U_n$ implies that $G(U_n)$ contains a proper division algebra of degree n. Hence $G(U_n)$ is a cyclic group of order n. Moreover, the structural theory of D_n yields that U_n must be a cyclic field. Thus we observe that all normal unramified fields U_n are cyclic, consequently all unramified fields are cyclic and there exists exactly one cyclic unramified extension Z_n of degree n over k. Or, the residue class field $\boldsymbol{k}$ possesses exactly one cyclic extension $\boldsymbol{Z}_n$ for each degree n.

Now it is very easy to prove that the local class field theory is valid in full extent for k if we assume furthermore that Theorem 3 holds also for all finite algebraic extensions K of k. According to the theory of C. Chevalley we can confine ourselves to show that each field K_n of degree n over k is splitting field of the cyclic group of all algebras of degree n for all unramified fields are cyclic.[14] Let $D = (\pi, Z_n, \sigma_n)$ be a generator of the group $G(Z_n)$. Then $D \times K_n \sim K_n$ for

$$\begin{aligned} D \times K_n &\sim (\pi, Z_nK_n/K_n, \Gamma(Z_nK_n, K_n)) \\ &\sim (\Pi^e, Z_nK_n/K_n, \sigma_e) \sim 1, \end{aligned}$$

where Π denotes a prime element of K_n and e, f denote the ramification and residue degrees of K_n respectively. Remark that $Z_n \cap K_n = Z_f$ is the inertial field of K_n as in the classical theory.[15]

THEOREM 2′. *If each extension* K_n *of degree* n *over* k *is splitting field of all algebras of degree* n *then there exists for each integer* n *exactly one cyclic unramified extension* Z_n *of degree* n *over* k.

Proof. For each integer n there exists at least one totally ramified field K_n, $\mathfrak{p} = \mathfrak{P}^n$ in K_n. Our assumption yields that K_n is splitting field of a division algebra D_n, hence the ramification degree e of D is equal to n. Consequently we can apply Theorem 2 together with Lemma 3 and see that the unramified extensions U_n of k are all cyclic.

THEOREM 3. *All unramified extensions of* k *are cyclic if the groups of*

---

[14] C. Chevalley, "La théorie du symbole de restes normiques," *Crelle*, vol. 169 (1932).

[15] Cf. Theorem 0 in the paper of Chevalley.

algebras $G(U_n)$ which belong to the unramified extensions U_n of degree n over k are cyclic groups of order n and if this property holds also for the respective groups over the finite algebraic extensions of the groundfield k.

Proof. According to the Galois theory it suffices to show that there exists for each degree n exactly one field U_n. Suppose then that U_n' and U_n'' are two distinct unramified normal fields of degree n then the relative degree $\bar{n}$ of their join $U_n'U_n''$ is greater than n. Hence $G(U_n'U_n'')$ is a cyclic group of order $\bar{n}$ containing the two cyclic groups $G(U_n')$ and $G(U_n'')$ both of order n. Consequently $G(U_n')$ and $G(U_n'')$ coincide. Now let D be an arbitrary normal division algebra of degree n lying in $G(U_n') = G(U_n'')$. The algebra D possesses U_n' and U_n'' as maximally commutative subfields. Our assumption implies in particular that there exist cyclic unramified extensions of degree n over k, namely otherwise we would arrive at a contradiction to Lemma 1. Consequently U_n' and U_n'' must be equal to a cyclic field if they are considered in the same algebraic closure of k.

In the preceding theorems we imposed conditions on the finite algebraic extensions of k as well as on the normal division algebras over k. We now wish to investigate the properties of perfect fields k for which only certain postulates concerning the normal division algebras are assumed, the conditions imposed upon the field k will be of purely commutative type.

THEOREM 4. *If all classes of normal algebras which are representable by division algebras of degree n form a group of order n and if for each prime q there exists at least one cyclic unramified extension Z_q of degree q over k then the cyclic unramified extensions of degree n are unique.*

Proof. It suffices to prove that there exists at least one cyclic unramified extension Z_n over k for each degree n. Namely, if Z_n is such an unramified field then

$$\begin{aligned}[G(Z_n):1] = [k^*/NZ^*_n:1] &= [\{\pi\}/\{\pi\}^n:1][\epsilon/NE:1] \\ &= n[\mathbf{k}^*/N\mathbf{Z}^*_n:1],\end{aligned}$$

and since $G(Z_n)$ is contained in the group of all algebras of degree n which is supposed to have the order n, we see that

> $G(Z_n)$ is a cyclic group of order n which coincides
> with the group of all algebras of degree n.

This statement is true for all cyclic unramified fields of degree n, moreover each generating algebra of the group of algebras of degree n is totally ramified and possesses all cyclic unramified fields of degree n as maximally commutative subfields. Hence these cyclic fields coincide in one and the same algebraic closure of k.

According to the Galois theory it is sufficient to show that there exist unramified cyclic fields Z_{q^ν} whose degrees are powers of a single prime q.

We distinguish two cases

$$\text{(i) } q = \chi(\boldsymbol{k}) \quad \text{and} \quad \text{(ii) } q \neq \chi(\boldsymbol{k}).$$

Case i. Since there exists at least one unramified cyclic field Z_q there exists at least one cyclic extension $\boldsymbol{Z}_q$ of degree q over the residue class field $\boldsymbol{k}$. According to the theory of cyclic extensions of degree q^ν over fields of characteristic q there consequently must exist such extensions $\boldsymbol{Z}_{q^\nu}$.[16] The corresponding perfect fields Z_{q^ν} are unramified and cyclic possessing the fields $\boldsymbol{Z}_{q^\nu}$ as residue class fields.[17]

Case ii. Again it is sufficient to consider the cyclic extensions of the residue class field $\boldsymbol{k}$. The field $\boldsymbol{k}$ contains a certain maximal q^N-th root of unity, where N is either a finite integer—then the q^{N+1}-st roots of unity do not lie in $\boldsymbol{k}$—or N is put equal to ∞ in a formal sense—then $\boldsymbol{k}$ contains all q^N-th roots of unity for arbitrarily great N. In the first case the $q^{N+\nu}$-th roots of unity determine cyclic extensions of degree q^ν for all such cyclotomic fields have a cyclic Galois group. In the second case our assumption yields too the existence of cyclic fields $\boldsymbol{Z}_{q^\nu}$ for any ν. Namely, it implies that the index $[\boldsymbol{k}^* : \boldsymbol{k}^{*q}]$ is divisible by q. Let $\boldsymbol{a}$ be a representative of $\boldsymbol{k}^*/\boldsymbol{k}^{*q}$ which is not equal to a q-th power, then the radicals $\boldsymbol{a}^{1/q^\nu}$ generate cyclic fields $\boldsymbol{Z}_{q^\nu} = \boldsymbol{k}(\boldsymbol{a}^{1/q^\nu})$ of degree q^ν.

Remark. The result of Theorem 4 can also be obtained if we substitute for the first assumption the following

> all classes of normal algebras over k possessing the
> exponent n form a set of n elements.

First we observe that all these classes form a group for if D_1 and D_2 are any two elements in the set, i. e. $D_1{}^n \sim D_2{}^n \sim k$, then

$$(D_1 \times D_2)^n \sim D_1{}^n \times D_2{}^n \sim k \quad \text{and} \quad (D_i{}^{-1})^n \sim (D_i{}^n)^{-1} \sim k.$$

---

[16] E. Witt, "Zyklische Körper und Algebren der Characteristik p vom Grade p^n," *Crelle*, vol. 176 (1937).

[17] H. Hasse, "Die Gruppe der p^n-primären Zahlen für einen Primteiler $\mathfrak{p}$ von p," *Crelle*, vol. 176 (1937).

Furthermore this group of algebras is cyclic; in order to prove this assertion it is sufficient to show that all groups of algebras possessing exponents q^i where q is a prime are cyclic. But this is obvious since the group having exponent q^i contains the group having exponent q^{i-1} and since the factor group has the order q according to our assumption. Now we may reason as before, the group of algebras of exponent q^i contains in particular the ramified division algebras (π, Z_{q^i}) belonging to the cyclic unramified fields Z_{q^i}.

We wish to observe that the unique existence of the cyclic unramified fields Z_n for each degree n over k does not imply that these fields are the only existing unramified extensions over k. Namely consider the following example. Let $\boldsymbol{k}_0$ be the field of all roots of unity over the field of all rational numbers; let $\boldsymbol{a}' \neq 0$ be an arbitrary element of $\boldsymbol{k}_0$ then there exists a maximal integer $N(\boldsymbol{a}')$ such that $\boldsymbol{a}'^{1/N(\boldsymbol{a}')}$ lies in $\boldsymbol{k}_0$ but not $\boldsymbol{a}'^{1/N(\boldsymbol{a}')\nu}$ for $\nu > 1$. Put $\boldsymbol{a} = \boldsymbol{a}'^{1/N(\boldsymbol{a}')}$. Next adjoin to $\boldsymbol{k}_0$ all solutions of all solvable equations whose normal fields do not possess the fields $\boldsymbol{k}_0(\boldsymbol{a}^{1/n})$, $n > 1$, as subfields. Let the resulting enumerable algebraic extension of $\boldsymbol{k}_0$ be $\boldsymbol{k}_1$. The field $\boldsymbol{k}_1$ is also quasi-algebraically closed, moreover $\boldsymbol{k}_1(\boldsymbol{a}^{1/n})$ are cyclic extensions of degree n over $\boldsymbol{k}_1$. Adjoin to $\boldsymbol{k}_1$ again all solutions of all solvable equations in $\boldsymbol{k}_1$ which do not contain $\boldsymbol{a}^{1/n}$, etc. The field $\boldsymbol{k}$ which we obtain finally is quasi-algebraically closed and it does not contain $\boldsymbol{a}^{1/n}$, moreover there do exist non-solvable extensions of sufficiently high degrees over $\boldsymbol{k}$. The perfect field k be now the field of all formal power series in one variable t with coefficients in $\boldsymbol{k}$. Evidentally all normal division algebras of degree n over $\boldsymbol{k}$ are powers of the algebra $(t, \boldsymbol{k}(\boldsymbol{a}^{1/n})\{t\}/k)$. Hence they form a cyclic group of classes of algebras having order n. The assumptions of Theorem 4 are fulfilled by the field k, but there exist other unramified extensions U_n over k which are not cyclic. Thus we see that the statement of Theorem 4 is the best possible.

In the following investigations we shall be concerned with the implications for the perfect field k if we assume certain properties to hold for the factor group of norm classes.

LEMMA 4. *If the class group k^*/NZ^*_n is cyclic of order n for all unramified cyclic extensions of degree n and if the same is true for all unramified extensions U of k then the field of residue classes $\boldsymbol{k}$ is quasi-algebraically closed.*

Proof. The assumption $k^*/NZ^*_n \cong \mathrm{Z}(n)$ yields that

$$\boldsymbol{k}^*/N\boldsymbol{Z}^*_n = 1,$$

for

$$[k^*/NZ^*_n : 1] = [\{\pi\}/\{\pi\}^n : 1][\epsilon/N\mathrm{E} : 1]$$
$$= n[\boldsymbol{k}^*/\boldsymbol{Z}^*_n : 1].$$

Moreover, we observe that the field of residue classes $\boldsymbol{k}$ is algebraically perfect. For otherwise there would exist proper cyclic and abelian division algebras $\boldsymbol{D}$ over $\boldsymbol{k}$ whose degrees are powers of the characteristic $\chi(\boldsymbol{k})$.[18] There would exist at least one cyclic unramified field Z_q over k whose field of residue classes is equal to a cyclic field $\boldsymbol{Z}_q$ which is splitting field of a normal division algebra over $\boldsymbol{k}$.[19] The norm class group k^*/NZ^*_q has then an order which is greater than q in contradiction to $k^*/NZ^*_q \cong Z(q)$.

Next let $\boldsymbol{D}$ be an arbitrary normal division algebra over $\boldsymbol{k}$, we must show that $\boldsymbol{D} \sim \boldsymbol{k}$. The algebra $\boldsymbol{D}$ possesses at least one normal splitting field $\boldsymbol{U}$: $\boldsymbol{D} \sim (\boldsymbol{a}(\sigma, \tau), \boldsymbol{U}, \Gamma(\boldsymbol{U}, \boldsymbol{k}))$. Let q be a prime divisor of the degree m of $\boldsymbol{D}$ which is supposed to be greater than 1. Σ_q be a Sylow subgroup of $\Gamma(\boldsymbol{U}, \boldsymbol{k})$ belonging to the prime q, and let $\boldsymbol{U}_q$ be the corresponding subfield of $\boldsymbol{U}$, then $[\boldsymbol{U}_q : \boldsymbol{k}] \not\equiv 0 \pmod{q}$. Since Σ_q is solvable there exist at least one chain of fields between $\boldsymbol{U}_q$ and $\boldsymbol{U}$

$$\boldsymbol{k} \leq \boldsymbol{U}_q = \boldsymbol{U}_q^{(0)} \leq \boldsymbol{U}_q^{(1)} \leq \cdots \leq \boldsymbol{U}_q^{(i-1)} \leq \boldsymbol{U}_q^{(i)} \leq \cdots \leq \boldsymbol{U}_q^{(s-1)} \leq \boldsymbol{U}_q^{(s)} = \boldsymbol{U}$$

such that $\boldsymbol{U}_q^{(i)}$ are cyclic extensions of degree q over $\boldsymbol{U}_q^{(i-1)}$, $i = 1, 2, \cdots, s$. Consider now the algebra $\boldsymbol{D} \times \boldsymbol{U}_q$, it will be similar to $\boldsymbol{U}_q$ as the argument used in the preceding Lemma 2 and the relation $\boldsymbol{k}^*/N\boldsymbol{Z}^*_n = 1$ which is supposed to hold also for the finite extensions of $\boldsymbol{k}$, show. Again $\boldsymbol{D} \times \boldsymbol{U}_q \sim \boldsymbol{U}_q$ leads to a contradiction. Hence $\boldsymbol{k}$ is quasi-algebraically closed, moreover it is algebraically perfect for fields $\boldsymbol{k}$ which are not algebraically perfect are centers of normal division algebras.

THEOREM 5. *If $k^*/NA^* \cong \Gamma(A, k)$ holds for all unramified abelian extensions A of k then there exists for each degree n a uniquely determined cyclic field of degree n.*

Proof. Let A be an arbitrary unramified abelian field of degree n over k, then

$$[k^*/NA^* : 1] = [\{\pi\}/\{\pi\}^n : 1][\epsilon/N\mathrm{E} : 1].$$

Hence

$$\Gamma(A, k) \cong Z(n) \quad \text{and} \quad \epsilon/N\mathrm{E} \cong \boldsymbol{k}^*/NA^* = 1,$$

i. e. all unramified abelian fields are cyclic. It is then obvious that there exists for each degree n exactly one cyclic unramified extension Z_n of k.

---

[18] A. A. Albert, "Normal division algebras over a modular field," *Transactions of the American Mathematical Society*, vol. 36 (1934); G. Köthe, "Über Schiefkörper mit Unterkörpern zweiter Art über dem Zentrum," *Crelle*, vol. 166 (1932).

[19] T. Nakayama, "Über die Algebren über einem Körper von der Primzahlcharakteristik, II," *Proceedings of the Imperial Academy of Tokyo*, vol. 12 (1937).

THEOREM 6. *If* $k^*/NU^* \cong \Gamma(A, k)$ *holds for all normal unramified extensions* U *over* k *where* A *denotes the maximal abelian subfield of* U *then there exist only cyclic unramified extensions* Z_n *over* k.

Proof. Let U_n be an arbitrary unramified normal extension of k then

$$[k^*/NU^*_n : 1] = [\{\pi\}/\{\pi\}^n : 1][\epsilon/N\mathrm{E} : 1] = n,$$

for

$$[k^*/NU^*_n : 1] = [k^*/NA^* : 1] \leqq n.$$

Consequently

$$\Gamma(A, k) \cong \mathrm{Z}(n),$$

hence

$$\Gamma(U_n, k) \cong \mathrm{Z}(n)$$

for $\Gamma(A, k)$ is a homomorphic map of $\Gamma(U_n, k)$.

Again as in Lemma 4 we see that $\boldsymbol{k}$ is algebraically perfect. Moreover, the field $\boldsymbol{k}$ is quasi-algebraically closed. Since $\boldsymbol{k}$ is algebraically perfect there do not exist normal division algebras of degree $\chi(\boldsymbol{k})^{\nu}$ over $\boldsymbol{k}$. Let then $\boldsymbol{D}$ be a normal division algebra over $\boldsymbol{k}$ whose degree is relatively prime to $\chi(\boldsymbol{k})$. Such an algebra is cyclic for $\boldsymbol{k}$ possesses only cyclic extensions $\boldsymbol{Z}_n$. The algebra $\boldsymbol{D}$ is the algebra of residue classes of a normal division algebra D over k belonging to $G(Z_n)$ where Z_n is the cyclic unramified extension of k corresponding to $\boldsymbol{Z}_n$. Since $G(Z_n)$ is a cyclic group of order n containing only ramified algebras the assumption $\boldsymbol{D} \not\sim \boldsymbol{k}$ leads to a contradiction.

LEMMA 5. *If* $[k^*/NZ^*_n : 1] = n$ *holds for all cyclic extensions of degree* n *over* k *then there exist cyclic unramified extensions of degree* n *with respect to* k.

Proof. We distinguish two cases

$$\text{(i)}\ \ (n, \chi(\boldsymbol{k})) = 1 \quad \text{and} \quad \text{(ii)}\ \ n = \chi(\boldsymbol{k})^{\nu}.$$

It is obviously sufficient to prove the assertion of the lemma for degrees n which are powers of a single prime q.

Case i. Suppose that there do not exist cyclic unramified extensions $Z_{q^{\nu}}$ of degree q^{ν}, then there do not exist cyclic extensions $\boldsymbol{Z}_{q^{\nu}}$ of degree q^{ν} over the field of residue classes $\boldsymbol{k}$. Hence the fields $\boldsymbol{k}$ and k respectively must contain all q^{r}-th roots of unity. Moreover, $[\boldsymbol{k}^* : \boldsymbol{k}^{*q}] = [\boldsymbol{k}^* : \boldsymbol{k}^{*q^{\nu}}] = 1$. For if $[\boldsymbol{k}^* : \boldsymbol{k}^{*q}]$ were divisible by q then any representative $\boldsymbol{a}$ of $\boldsymbol{k}^*/\boldsymbol{k}^{*q}$ which is not equivalent to 1 mod $\boldsymbol{k}^{*q}$ would generate cyclic extensions $\boldsymbol{Z}_{q^{\nu}} = \boldsymbol{k}(\boldsymbol{a}^{1/q^{\nu}})$ of degree q^ν over $\boldsymbol{k}$. Since $[k^*/NZ^*_n:1]=n$ for all unramified fields Z_n implies that $\boldsymbol{k}$ is algebraically perfect there is a one to one correspondence between the cyclic unramified extensions Z_{q^ν} over k and the cyclic extensions $\boldsymbol{Z}_{q^\nu}$ over $\boldsymbol{k}$, the field $\boldsymbol{Z}_{q^\nu}=\boldsymbol{k}(\boldsymbol{a}^{1/q^\nu})$ must be equal to $\boldsymbol{k}$.

Next $[\boldsymbol{k}^*:\boldsymbol{k}^{*q^\nu}]=1$ implies that all elements of $\boldsymbol{k}$ are q^ν-th powers. Consequently all units ϵ of k are q^ν-th powers as a simple p-adic approximation yields.

Now consider the cyclic ramified field $K_{q^\nu}=k(\pi^{1/q^\nu})=k((\epsilon\pi)^{1/q^\nu})$. The group of all norms of elements in K_{q^ν} contains the group of all units ϵ of k for they are all q^ν-th powers as we have just observed. Moreover $NK^*_{q^\nu}$ contains the prime element π, namely if $q\neq 2$ then $N(\pi^{1/q^\nu})=\pi$ and if $q=2$ then -1 is a square or 2^ν-th power in k consequently $\pi=-1\cdot-\pi$ is a norm. Hence $NK^*_{q^\nu}=k^*$ in contradiction to the assumption of the Lemma. Thus we find that k must possess cyclic unramified extensions Z_{q^ν}, they will be found among the cyclotomic extensions if ζ_{q^N} (N finite) is the maximal root of unity lying in k or they will be radical fields if all q^ν-th roots of unity lie in k.

Case ii. The assumption yields that there must exist proper division algebras D of degree q^ν over k, namely among possible other algebras the generators D_{q^ν} of the cyclic groups $G(Z'_{q^\nu})$ where Z'_{q^ν} denotes a ramified cyclic extension of degree q^ν over k. The existence of such fields has been established in the general theory of cyclic extensions over perfect fields whose residue class fields $\boldsymbol{k}$ are algebraically perfect.[20]

Now let D_{q^ν} be an arbitrary division algebra of degree q^ν over k, it possesses at least one unramified normal splitting field U for $\boldsymbol{k}$ is algebraically perfect:

$$D_{q^\nu}\sim(a(\sigma,\tau),U,\Gamma(U,k))\sim(\pi^{c(\sigma,\tau)},U,\Gamma(U,k))(\eta(\sigma,\tau),U,\Gamma(U,k)).$$

The algebra $(\eta(\sigma,\tau),U,\Gamma(U,k))$ which could be similar to an unramified algebra of degree q^μ $(\mu\leq\nu)$ must be similar to k for its residue algebra $(\eta(\sigma,\tau) \bmod \mathfrak{p},\boldsymbol{U},\Gamma(\boldsymbol{U},\boldsymbol{k}))$ is similar to $\boldsymbol{k}$, $\boldsymbol{k}$ being an algebraically perfect field.[21] Hence

$$D_{q^\nu}\sim(\pi^{e(\sigma,\tau)},U,\Gamma(U,k))\sim(\pi,Z_{q^\rho})^\sigma,$$

where Z_{q^ρ} denotes the unramified cyclic subfield of U which corresponds to the character of $\Gamma(U,k)$ induced by the set of addends $e(\sigma,\tau)$.

---

[20] Cf. the two papers mentioned in footnotes 16 and 17.

[21] E. Witt, "Schiefkörper über diskret bewerteten Körpern," *Crelle*, vol. 176 (1937).

We have $\rho = \nu$ for D_{q^ν} was supposed to be a normal division algebra of degree q^ν, thus the existence of cyclic unramified fields of degree q^ν is established. Moreover we observe that all algebras D_{q^ν} are cyclic and ramified.

THEOREM 7. *If* $[k^*/NZ^*_n : 1] = n$ *holds for all cyclic extensions* Z_n *over* k *and similarly for all cyclic extensions over the finite algebraic extensions of* k *then*

(i) *all cyclic unramified fields of degree* n *over* k *coincide,*
(ii) *the field of residue classes* $\boldsymbol{k}$ *is algebraically perfect, and*
(iii) $k^*/NA^* \cong \Gamma(A, k)$ *holds for all abelian extensions* A *over* k.

Proof. First we observe that Lemma 5 implies the existence of cyclic unramified fields Z_q of prime degree q. If $q \neq \chi(\boldsymbol{k})$ then there exist cyclic unramified extensions Z_{q^ν} for arbitrary ν. Namely, either cyclotomic subfields of $k(\zeta_{q^{N+\nu}})$—if ζ_{q^N} is the maximal q^N-th root of unity lying in k—represent such fields or radical extensions $k(a^{1/q^\nu})$ where $Z_q = k(a^{1/q})$ is one of the cyclic unramified extensions which must exist according to Lemma 5. If $q = \chi(\boldsymbol{k})$ then cyclic unramified extensions must exist according to the theory of q-algebras if the latter is applied to perfect fields for which our assumption holds.

Lemma 5 implies furthermore that $\boldsymbol{k}$ is algebraically perfect. Hence there can be established a one to one correspondence between the unramified division algebras over k and the division algebras over $\boldsymbol{k}$.

In order to prove that all cyclic unramified fields of degree n coincide it is sufficient to prove that all cyclic unramified fields of prime degree q^ν coincide for each q. Let then Z_q and Z'_q be two cyclic unramified fields of degree q over k. Their respective groups of algebras $G(Z_q)$ and $G(Z'_q)$ have the generators (π, Z_q) and (π, Z'_q) respectively. Consequently—if $q \neq \chi(k)$—the separable field $k(\pi^{1/q})$ is a common splitting field of $G(Z_q)$ and $G(Z'_q)$. Now assume that k contains the q-th roots of unity then $k(\pi^{1/q})$ is a cyclic extension of degree q over k. Our assumption implies

$$G(Z_q) = G(Z'_q) = G(k(\pi^{1/q})) \cong \mathrm{Z}(q).$$

Since each proper division algebra D_q in $G(k(\pi^{1/q}))$ is ramified and since Z_q and Z'_q are both unramified cyclic splitting fields of D_q we get for the algebra of residue classes

$$\mathfrak{O}/\mathfrak{P} = \mathfrak{O}(Z_q)/\mathfrak{P} = \mathfrak{O}(Z'_q)/\mathfrak{P}, \text{ or } \boldsymbol{Z}_q = \boldsymbol{Z}_q' \text{ over } \boldsymbol{k},$$

hence Z_q and Z'_q are abstractly isomorphic which amounts exactly to our assertion.

If k does not contain the q-th roots of unity then consider the extensions of $G(Z_q)$ and $G(Z'_q)$ by the cyclotomic field $k(\zeta_q)$. Since $[k(\zeta_q):k]$ is relatively prime to q the general algebraic theory of splitting fields yields that the algebras in $G(Z_q)$ and $G(Z'_q)$ to not become similar to $k(\zeta_q)$ if the center is extended to $k(\zeta_q)$. The field $k(\pi^{1/q})k(\zeta_q)$ is a cyclic extension of degree q over the new ground field $k(\zeta_q)$. As in the previous case our assumption yields that the extended groups coincide. Hence the fields $Z_qk(\zeta_q)$ and $Z'_qk(\zeta_q)$ which are cyclic of degree $q(q-1)$ over k coincide, consequently according to the Galois theory $Z_q = Z'_q$.

A treatment of the case $q = \chi(\boldsymbol{k}) = \chi(k)$ can be found in the literature.[22]

Next we show that the field of residue classes $\boldsymbol{k}$ is quasi-algebraically closed. Our assumption implies in particular that $k^*/NZ^*_n \cong \mathrm{Z}(n)$ for all cyclic unramified extensions of degree n. Consequently there do not exist cyclic algebras over $\boldsymbol{k}$ for such algebras would be residue algebras of unramified cyclic algebras over k. Furthermore our general assumption implies the same for all finite algebraic extensions over $\boldsymbol{k}$. Hence we can apply Lemma 2 and we see that $\boldsymbol{k}$ is quasi-algebraically closed.

An immediate consequence according to Theorem 1 and the algebraic theory is the fact that all normal division algebras of degree n over k are ramified and that they possess cyclic unramified splitting fields Z_n, moreover degree and exponent of all division algebras coincide. Since the fields Z_n are uniquely determined by their degrees n we see that all algebras which are representable by division algebras of degree n form a cyclic group of order n.

Next we prove that all abelian extensions A_n of degree n over k are splitting fields of the class group of all algebras of degree n over k. Let Z_f be the inertial field of a fixed abelian field A_n, then $\mathfrak{p} = \mathfrak{P}^{n/f}$ in A_n and $\pi = \mathrm{E}\Pi^{n/f}$ with a suitable unit E of A_n. Consider now a generating algebra (π, Z_n, σ_n) of $G(Z_n)$ and form the direct product $(\pi, Z_n, \sigma_n) \times A_n$ then

$$\begin{aligned}(\pi, Z_n, \sigma_n) \times A_n &\sim (\Pi^{n/f}, Z_nA_n/A_n, \sigma_{n/f})(\mathrm{E}, Z_nA_n/A_n, \sigma_{n/f})\\ &\sim (\Pi, Z_nA_n/A_n, \sigma_{n/f})^{n/f} \sim 1\end{aligned}$$

for our assumptions imply that $A^*_n/N(Z_nA_n)^* \cong \mathrm{Z}(n/f)$ and Z_nA_n is the unramified cyclic extension of degree n/f over A_n. Hence A_n is a splitting field of the uniquely determined class group of algebras $G(Z_n)$.

Now we prove that $[k^*/NS^*:1] \leqq [S:k]$ for any solvable extension S of k. Since S is solvable it possesses a chain of subfields

$$k = A_0 \leqq A_1 \leqq \cdots \leqq A_{i-1} \leqq A_i \leqq \cdots \leqq A_{s-1} \leqq A_s = S$$

---

[22] See footnote 20; consider the algebras (a, π^{-1}) which have the same unramified splitting field.

such that always A_i is a cyclic extension over A_{i-1}. The cyclicity of $G(Z_n)$ yields that the norm factor group for any cyclic extension over k is cyclic, according to our general assumption the same is true for all finite algebraic extensions K over k. Thus $[A^*_{i-1} : NA^*_i] = [A_i : A_{i-1}]$. Using the transitive property of the norm and induction by i one readily observes that $[k^* : NS^*] \leqq [S : k]$.[23]

Now let A_n be an Abelian field over k, there always exists a division algebra $D_n = (a(\sigma, \tau), A_n, \Gamma(A_n, k))$ in $G(Z_n)$ as we have seen before. Hence the function $f(\sigma) = \prod_{(\tau)} a(\sigma, \tau)$ determines an isomorphism between $\Gamma(A_n, k)$ and a subgroup of k^*/NA^*_n, consequently

$$k^*/NA^*_n \cong \Gamma(A_n, k)$$

as asserted.[24]

Remark. Our assumptions do not exclude the existence of non-solvable equations in k as the example on page 86 shows.

The uniqueness of the cyclic unramified extensions in Theorem 6 was a consequence of the postulated isomorphism $k^*/NA^* \cong \Gamma(A, k)$. This isomorphism amounts to the fact that the norm factor group of a cyclic unramified extension over another cyclic unramified extension over the ground field k is a cyclic group whose order is equal to the respective relative degree. The same result as in Theorem 6 can be obtained if we postulate for k

(i) $k^*/NZ^*_n \cong Z(n)$ for all cyclic unramified fields,

(ii) if Z_n and Z_m are two cyclic fields then

$$NZ^*_n \cap NZ^*_m = N(Z_n Z_m)^*, \quad \text{or}$$

(ii′) $\{NZ^*_n, NZ^*_m\} = N(Z_n \cap Z_m)^*$.

Obviously it suffices to prove the uniqueness for cyclic unramified fields Z'_{q^ν} and Z''_{q^μ} whose degrees are powers of one and the same prime q. Assume that Z'_{q^ν} and Z''_{q^μ} are distinct. Then

$$q^M < [Z'_{q^\nu} Z''_{q^\mu} : k] \leq q^{\mu+\nu} \quad \text{and} \quad q^m > [Z'_{q^\nu} \cap Z''_{q^\mu} : k] \geq 1.$$

---

[23] For a model of the proof see F. K. Schmidt, "Zur Klassenkörpertheorie im Kleinen," *Crelle*, vol. 162 (1930).

[24] T. Nakayama, "Über die Beziehungen zwischen den Faktorensystemen und der Normenklassengruppe eines galoisschen Erweiterungskörpers," *Mathematische Annalen*, vol. 112 (1935).

Since both fields are unramified the group of units ϵ in k is contained in $NZ_{q^{\nu}}'^{*}$ as well as in $NZ_{q^{\mu}}''^{*}$ assumptions (ii) and (ii′) together with (i) yield

$$N(Z_{q^{\nu}}'Z_{q^{\mu}}'')^{*} = q^{M} \quad \text{and} \quad N(Z_{q^{\nu}}' \cap Z_{q^{\mu}}'')^{*} = q^{m},$$

$M = \max(\mu, \nu)$ and $m = \min(\mu, \nu)$. But surely $[k^{*}/N(Z_{q^{\mu}}'Z_{q^{\nu}}'')^{*} : 1] > q^{M}$ and $[k^{*}/N(Z_{q^{\mu}}' \cap Z_{q^{\nu}}'')^{*} : 1] < q^{m}$. Hence $Z_{q^{m}}$ must be a subfield of $Z_{q^{M}}$ whereby the uniqueness is established.

Lemma 6. *If for each subgroup H of the multiplicative group k^{*} belonging to the perfect field k there exists a uniquely determined abelian field K of finite degree over k such that $NK^{*} = H(K) = H$ and if $H_1 < H_2$ implies $K_2 < K_1$ then*

$$[k^{*}/H : 1] = [K : k].$$

Proof. Let H be an arbitrary subgroup of k^{*} and let K be the associated abelian field over k such that $H(K) = H$. The Galois group of K with respect to k be denoted by $\Gamma(K, k) = \Gamma$. Consider now an arbitrary chain of subfields between K and k

$$k = K_0 < K_1 < K_2 < \cdots < K_{i-1} < K_i < \cdots < K_{s-1} < K_s = K$$

such that K_i is a cyclic extension K_{i-1} whose degree is a prime, $i = 1, 2, \cdots, s$. The associated norm groups $N_i K^{*}_{i} = H(K_i) = H_i$ form a descending chain

$$k = H_0 \geq H_1 \geq \cdots \geq H_{i-1} \geq H_i \geq \cdots \geq H_{s-1} \geq H_s = H$$

as a simple consideration shows. According to our assumption there holds a one to one correspondence between the subgroups H_i and the subfields K_i, K_i being the uniquely determined abelian extension of k belonging to $H_i = N_i K^{*}_{i}$.

Always $H_{i-1} > H_i$, for if $H_{i-1} = H_i$ then also $K_{i-1} = K_i$ according to the second part of our assumption. Moreover, the factor groups H_{i-1}/H_i are cyclic groups of prime order for the existence of a proper subgroup H' of H_{i-1} which is different from H_i would imply the existence of a field K' such that $K_{i-1} < K' < K_i$.

Hence $[k^{*}/H : 1] = \prod_{i=1}^{s} [H_{i-1}/H_i : 1]$ is finite.

Now let H be a subgroup of k^{*} such that $[k^{*}/H : 1] = q^{\rho}$ where q is a rational prime. We draw a composition series

$$k = H_0 > H_1 > H_2 > \cdots > H_{i-1} > H_i > \cdots H_{\rho-1} > H_{\rho} = H$$

between k^* and H such that the factor groups H_{i-1}/H_i are cyclic groups of order q. The associated abelian extensions of k be

$$k = K_0 < K_1 < K_2 < \cdots < K_{i-1} < K_i < \cdots < K_{\rho-1} < K_\rho = K,$$

evidentally K_i is always a cyclic extension of prime degree q_i over K_{i-1}, $i = 1, 2, \cdots$. Next we show that $q_i = q$. Consider the field K_1 over k then $k > H_1 > k^{*q_1}$, since $[k^* : H_1] = q$ we get $q_1 = q$. Then induction implies $q_2 = \cdots = q_\rho = q$. Hence obviously

$$[k^*/H : 1] = [K : k].$$

Now let H be a subgroup of index $n = \prod_{i=1}^{r} q_i^{\rho_i}$ under k. According to a group theoretical theorem we obtain $H = H^{(1)} \cap \cdots \cap H^{(r)}$ where $H^{(i)}$ are subgroups of index $q_i^{\rho_i}$ under k^*. The Galois theory combined with the assumptions of the Lemma yields that the field K belonging to H is the join of the fields $K^{(i)}$ belonging to the groups $H^{(i)}$. Hence $[k^*/H : 1] = [K : k]$.

COROLLARY. *For each finite abelian extension K over k holds*

$$[k^*/H(K) : 1] = [K : k].$$

Proof. We draw again a composition chain between k and $H(K)$. The finiteness of the degree $[K : k]$ implies according to the assumptions of the lemma that the index $[k^*/H(K) : 1]$ is finite. The uniqueness of K as the field belonging to $H(K)$ asserts together with Lemma 6 that

$$[k^*/H(K) : 1] = [K : k].$$

THEOREM 8. *Under the same assumptions as in Lemma 6 it follows that all abelian unramified fields U_n are cyclic.*

Proof. Let U_n be an arbitrary abelian unramified field of degree n over k. Then $H(U_n) = (\{\pi\}^n, NE)$. Since $[k^* : H(U_n)] = n$ according to the corollary we must have $NE = \epsilon$. The last equality holds for all abelian unramified extensions U_n, hence their norm groups coincide and they must be cyclic for this conclusion holds in particular for all prime degrees q.

THEOREM 9. *If the abelian field K over k belonging to the group H is contained in all extensions L over k whose norm groups $H(L)$ are subgroups* *of H and if the assumptions of Lemma 6 hold for k then all unramified finite extensions U_n of k are cyclic.*

Proof. Let U_n be an unramified separable field of degree n over k then $H(U_n) = (\{\pi\}^n, NE)$. The group NE is a subgroup of ϵ the group of all units in k, hence $H(U_n) \subseteq (\{\pi\}^n, \epsilon)$. Let the cyclic unramified extension belonging to $(\{\pi\}^n, \epsilon)$ according to Theorem 8 be Z_n. Then our assumption yields that $Z_n \subseteq U_n$. Consequently $Z_n = U_n$ for both fields have the same degree.

Remark. A further consequence of the assumptions in Lemma 6 is the fact that the field of residue classes $\boldsymbol{k}$ is algebraically perfect. Namely Theorem 8 yields that $k^*/NZ^*_n \cong Z(n)$ holds for all unramified extensions, hence there do not exist proper division algebras of degree $\chi(\boldsymbol{k})^\nu$ over $\boldsymbol{k}$ as previous considerations show.

The most important conclusion obtained in the previous theorems was the fact that for each integer n there exists a uniquely determined cyclic unramified extension Z_n of degree n over k. In some theorems we had to assume that the field of residue classes be an algebraically perfect field mostly then if we made use of the algebraic theory of division algebras D over k in the proofs. However, in theorems of this type it suffices in general that the assumptions are postulated for all normal division algebras which possess residue division algebras whose centers are separable over $\boldsymbol{k}$. Our aim is to infer from the postulates that $\boldsymbol{k}$ is a Galois field. In order to achieve this it is necessary to impose further conditions on the perfect field k. Namely, there exist examples of perfect fields for which most of the theorems hold although the residue fields are not Galois fields. For example the fields of formal power series in two variables over an algebraically closed field of characteristic 0 where the residue field belonging to only existing isolated subgroup of rank one contained in the valuation group of the field is considered as the residue field of the discrete archimedian valuation of the field.[25]

The restriction to be imposed on k should not contain a postulate with regard to the characteristic of $\boldsymbol{k}$. Thus we are lead to the following assumption *the postulates in the theorems asserting the uniqueness of the unramified cyclic extensions over k shall also be true for all perfect subfields k' of k whose residue class fields $\boldsymbol{k}'$ are subfields of $\boldsymbol{k}$.*

THEOREM 10. *If a field k satisfying any one of the assumptions implying*

---

[25] O. F. G. Schilling, "Arithmetic in fields of formal power series," *Annals of Mathematics*, vol. 38 (1937).

the uniqueness of the cyclic unramified extensions fulfills the postulate about its perfect subfields then k is a perfect field whose field of residue classes $\boldsymbol{k}$ is a Galois field whose G-number possesses no infinite part.[26]

Proof. The field of residue classes $\boldsymbol{k}$ must have the property that all its subfields $\boldsymbol{k}'$ admit exactly one cyclic extension of degree n, hence in particular the prime field of characteristic $\chi(\boldsymbol{k})$ which lies in $\boldsymbol{k}$. Thus fields $\boldsymbol{k}$ of characteristic 0 are ruled out for the field of all rational numbers possesses infinitely many cyclic extensions for each degree n. For the same reason it follows that $\boldsymbol{k}$ must be algebraic over its prime field if $\chi(\boldsymbol{k}) \neq 0$. Namely, if $\boldsymbol{k}$ would contain a transcendental quantity t over the prime field then the field of all rational functions of t over the prime field had to possess exactly one cyclic extension of degree n for each integer n, an implication which obviously is false.

Now it remains to be proved that the infinite part of the G-number belonging to $\boldsymbol{k}$ is equal to one. Assume that q is a divisor of $G_{\text{inf}}(\boldsymbol{k})$. Then $\boldsymbol{k}$ possesses no cyclic extensions of degree q^{ν} over $\boldsymbol{k}$, consequently the assumptions which lead to the uniqueness of the cyclic unramified extensions over k are violated for all degrees which are divisible by q. An extensive investigation leading up to the preceding assertion has already been made.[27] Hence $G_{\text{inf}}(\boldsymbol{k}) = 1$.

We see that the residue fields $\boldsymbol{k}$ which result are algebraically perfect. Therefore the structure of the perfect fields k is determined by the fields $\boldsymbol{k}$.[28] In a certain sense they are the $\mathfrak{p}$-adic fields belonging to algebraic number fields and fields of functions of one variable whose fields of coefficients are Galois fields; these are exactly the fields for which the local class field theory has first been developed. We wish to mention that the perfect fields which can be characterized by topological properties belong to the class of perfect fields we studied here.[29] Thus we have studied an algebraico-arithmetical counterpart of N. Jacobson's theory.

It was quite material in our investigations that the value group $\mathfrak{V}(k)$ belonging to the perfect field was isomorphic with the additive group of all rational integers. It implies in particular that the prime ideal $\mathfrak{p}$ of k can be generated by a prime element π. Consequently we could speak of a ramification degree e belonging to a finite algebraic extension of k or to a normal division algebra over k. However, one can find a substitute for this definition of ramification. Namely, if $e > 1$ then $[\mathfrak{O}/\mathfrak{P} : \mathfrak{o}/\mathfrak{p}] = n/e < n$ in the case that $\mathfrak{V}(k)$ is isomorphic to the additive group of all integers, if $\mathfrak{V}(k)$ is not discrete then we shall say that K is ramified over k if $[\mathfrak{O}/\mathfrak{P} : \mathfrak{o}/\mathfrak{p}] < n = [K : k]$. One easily observes that this is a workable definition for arithmetical problems.

---

[26] For the pertaining notations and definitions see the paper mentioned in footnote 1.

[27] See paper mentioned under [1] and an additional note in the same journal.

[28] H. Hasse and F. K. Schmidt, "Die Struktur diskret bewerteter Körper," *Crelle*, vol. 170 (1933).

[29] N. Jacobson, "Totally disconnected locally compact rings," *American Journal of Mathematics*, vol. 58 (1936).

More important are the consequences of the non-discreteness of $\mathfrak{V}(k)$ for the algebraic theory of normal division algebras. We made ample use of the fact that each division algebra which possesses a residue algebra whose center is separable over $\boldsymbol{k}$ can be represented as the product of a ramified and an unramified algebra

$$\begin{aligned} D &\sim (a(\sigma,\tau), U, \Gamma(U,k)) \\ &\sim (\pi^{e(\sigma,\tau)}, U, \Gamma(U,k))\,(\eta(\sigma,\tau), U, \Gamma(U,k)). \end{aligned}$$

For general value groups $\mathfrak{V}(k)$ of rank one such a decomposition is not obvious for there do not exist generating prime elements of the prime ideal belonging to the valuation $\mathfrak{V}(k)$. Nevertheless, under restricting conditions on the structure of the field k we are able to find a substitute for such decompositions; but it turns out that such decompositions can no more be made unique as it was possible in the case of discrete valuations by fixing the prime element π. Let us assume that there exists at least one discrete perfect subfield k' of k which has the same field of residue classes $\boldsymbol{k}$ as the given field k, and that there exists a—necessarily infinite—algebraic extension k'' of k' which is everywhere dense in k, that is to say whose derived field with respect to the valuation is equal to k. The values assumed by the elements of k'' form a group of rational numbers $\mathfrak{V}_e(k'')$ whose structure is determined by the G-degree of k'' over k'. In order to see that one has to observe that any number of k'' lies in a finite perfect extension of k' and that its value is determined by the relative degree of that extension and that the value does not depend upon the particular finite field used for its determination. Analyzing the structure of the group $\mathfrak{V}_e(k'')$ we see that certain elements of it may possess arbitrarily high powers of primes as denominators. The set of such rational primes shall be called the infinite characteristic of $\mathfrak{V}_e(k'')$, it is determined by the infinite part of the G-degree $[k'' : k']$; we collect all these primes q in a formal product $\prod_{(q)} q = G$.

Now it can be shown that each algebra D over k whose algebra of residue classes has a separable center over $\boldsymbol{k}$ is equivalent to the direct product of a suitable algebra D'' over k'' with k.[30] Thus the study of normal algebras over k is reduced to the investigation of the normal algebras over k''.

Now let D'' be an arbitrary division algebra over k'',

$$D'' \sim (a(\sigma,\tau)'', U'', \Gamma(U'', k'')),$$
$$U'' = k''(A'') \text{ where } A''^n + a''_1 A''^{n-1} + \cdots + a''_n = 0 \text{ with } a''_i \text{ in } k''.$$

Consider now an arbitrary finite algebraic extension k_* of k' which contains the elements $a(\sigma,\tau)''$ and a''_i. Then $U'' = k_*(A'')k''$ and

$$D'' \sim (a(\sigma,\tau)'', k_*(A'')/k_*, \Gamma) \times k'' \sim D_* \times k''.$$

Since k_* is a finite algebraic extension of the discrete perfect field k' there exists a prime element π_* in k_*, hence

$$D_* \sim (\pi_*^{c(\sigma,\tau)}, k_*(A'')/k_*, \Gamma)(\eta(\sigma,\tau)_*, k_*(A'')/k_*, \Gamma),$$

where the first factor represents a ramified algebra and the second factor stands for an unramified algebra. The algebras

$$(\pi_*^{c(\sigma,\tau)}, k_*(A'')/k_*, \Gamma) \times k'' \quad \text{and}$$
$$(\eta(\sigma,\tau)_*, k_*(A'')/k_*, \Gamma) \times k''$$

are at most similar to a ramified or an unramified division algebra over k'', respectively, where the ramification is understood to be measured according to the new definition. The resulting decomposition of D'' depends essentially upon the choice of the subfield k_*, nevertheless the distinction between the ramified and unramified part still persists.

---

[30] Cf. Note 1. The result essentially used here can be generalized as follows: "If k'' is an everywhere dense subfield of k such that the residue class fields of k'' and k coincide, every separable extension $k(A)$ of degree n over k is equal to the join of k with a suitable separable extension $k''(A'')$ of degree n over k''." A proof of this theorem can be obtained by generalizing a proof of M. Moriya in "Klassenkörpertheorie im Kleinen fur die unendlichen algebraischen Zahlkörper" (*Journal of the Fac. of Sci. Hokkaido Imp. Univer.*, ser. I, vol. 5 (1936), Sapporo, Japan). Using the notations *loc. cit.*, p. 13, we observe that one can take instead of k any everywhere dense subfield k'' of k. Furthermore, in the general case one has to use for the construction of the equation for A'' the theory of abstract derivation as developed by H. Hasse in "Noch eine Begründung der Theorie der höheren Differentialquotienten in einem algebraischen Funktionenkörper einen Unbestimmten" (*Crelle*, vol. 177 (1937). The proof that the constructed equation for A'' is irreducible in k'' resp. k can be reduced to a theorem of F. K. Schmidt stating that two polynomials of degree n whose distance is sufficiently small in the metric space of all polynomials of degree n, have the same type of decomposition. Cf. F. K. Schmidt, "Mehrfach perfekte Körper" in *Mathematische Annalen*, vol. 108 (1933). Our theorem is rather helpful for the investigation of the structural theory of general perfect fields whose value groups are isomorphic with non-discrete subgroups the additive group of all real numbers.

$$(\pi_*^{e(\sigma,\tau)}, k_*(A'')/k_*, \Gamma) \times k'' \sim (\pi_*^{\nu}, Z_*/k_*, \sigma) \times k'',$$

where Z_* denotes the cyclic unramified subfield of $k_*(A'')$ belonging to the character induced by $e(\sigma,\tau)$. The cyclic field $Z'' = Z_* k''$ can also be described as the subfield of U'' which is determined by the abelian character of Γ singled out by the values of the original factor set $a(\sigma,\tau)''$ within $\mathfrak{V}(k)$ or $\mathfrak{V}(k'')$ for the exponents $e(\sigma,\tau)$ are uniquely determined modulo $[U'':k]$ within $\mathfrak{V}(k_*)$ or $\mathfrak{V}(k'')$.

If the prime r does not divide G, then we can select the field k_* used for the construction of a division algebra D'' whose degree is a power of r such that the relative degrees of all fields between k_* and k'' are relatively prime to r. Hence k'' does not split the related algebra $D_* \not\sim k_*$, of course provided that such an algebra exists.[31] In such a case the algebra D'' is represented as the direct product of a ramified and an unramified algebra as we observed before.

Now let q be a divisor of G, and let $D_* = (\pi_*, Z_{q^\nu}, \sigma_{q^\nu})$ be a division algebra over k_*. Here we can no more assume that the relative degrees of all fields between k_* and k'' are relatively prime to q. Let K be a sufficiently large extension of k_* such that $[K:k_*] = q^\lambda s$ is divisible by q^ν. Then

$$D_* \times K \sim (\Pi^{q^\lambda s}\mathrm{E}, Z_{q^\nu}K/K, \sigma_{q^\nu}) \sim (\mathrm{E}, Z_{q^\nu}K/K, \sigma_{q^\rho}),$$

where Π denotes a prime element of the field K and where $[Z_{q^\nu}K:K] = q^\rho$. Hence the algebra $D_* \times K$ and a fortiori the algebra $D_* \times k''$ are equivalent to unramified algebras over K and k'' respectively. Thus we see that there do not exist ramified division algebras D'' over k'' whose degree is a power of a prime q dividing the infinite characteristic G. Observe that there may very well exist unramified division algebras over k'' whose degree is a power of such a prime q.

We next wish to construct an example of a perfect field k whose value group $\mathfrak{V}(k)$ is not discrete and where all normal algebras of degree n form a cyclic group of order n. Let $\boldsymbol{k}$ be a finite p-adic number field, then all classes of algebras which can be represented by normal division algebras whose degrees are divisors of n form a cyclic group of order n for each integer n. Consider the field $k' = \boldsymbol{k}\{t\}$ of all formal power series in t which possess coefficients in the field $\boldsymbol{k}$. The adjoin to k' all radicals t^{1/q^ν} where q runs over all rational primes and $\nu = 1, 2, \cdots$. The resulting enumerable infinite extension k'' possesses an infinite characteristic G which is divisible by all rational primes. According to what we have seen before all division algebras of degree n are unramified and consequently they are determined by the division algebras $\boldsymbol{D}$ over $\boldsymbol{k}$, since $\boldsymbol{k}$ is algebraically perfect there can be established a one to one correspondence between the algebras over k'' and $\boldsymbol{k}$. Thus we see that all classes of algebras over k'' which can be represented by division algebras whose degrees are divisors of n for a cyclic group of order n. The same is true for the classes of normal algebras over the derived field k belonging to k''. One readily observes that the local class field theory in the form of the isomorphism theorem stating the isomorphism of the Galois group of an abelian extension A over k with the respective norm class group in k does not hold. Thus we observe that the discreteness of the value group $\mathfrak{V}(k)$ is essential for the validity of all theorems known in the usual local class field theory.

---

[31] For a more detailed treatment of the perfect fields arising from infinite algebraic number fields see [1].

As a matter of fact one readily can construct examples of fields k which are perfect with respect to a non-discrete valuation of rank one—or of fields which are everywhere dense in such fields—such that certain algebras are ramified and others are unramified according to the respective degrees. One has to consider suitable infinite perfect fields $\boldsymbol{k}$ for which the local class field theory holds in part, i. e. such that all algebras whose degrees divide a fixed G-number form cyclic groups, then one adjoins to $\boldsymbol{k}\{t\}$ sufficiently many algebraic quantities such that the infinite characteristic of the resulting field k'' is equal to the previously fixed G-number and such that the algebras over k'' whose degrees are relatively prime to the G-number form cyclic groups. The resulting field k'' has the property that all algebras whose degrees divide a fixed integer n form a cyclic group of order n. Moreover, examples of such a type show that the cyclicity of class groups of algebras over a field does not imply that the field is perfect. Repeating the process of adjoining transcendental quantities like t one can find fields which admit a non-discrete valuation of arbitrary type of ordering but such that the special property of the class groups of algebras holds.

THE JOHNS HOPKINS UNIVERSITY.

GROUPS WHOSE COMMUTATOR SUBGROUPS ARE OF ORDER TWO.*

By G. A. Miller.

If the commutator subgroup of a group G is of order two the commutators of G are invariant and hence every operator of odd order appears in the central of G since such an operator could not be transformed into an operator of even order. It therefore results that when G involves operators of odd order it is the direct product of a group of order 2^m and of an abelian group of odd order. It is desirable to exclude direct products in what follows since all except one of the factor groups would be abelian. Hence we shall assume hereafter that the order of G is 2^m and that G is not a direct product. The central of G includes the squares of all the operators of G and hence the central quotient group of G is abelian and of type 1^n, where n is even since the subgroup composed of all of the operators of G which are commutative with two of its non-commutative operators is of index 4 under G.

It is possible to construct as follows a G whose central is an arbitrary abelian group. If $t_1, t_2, \cdots, t_l$ is a set of restricted independent generators of this abelian group we divide these operators into distinct pairs when l is even or we divide $l-1$ of them into distinct pairs when l is odd. In the former case we construct two operators whose squares are the operators of such a pair and that one of these two operators generates the commutator subgroup. Each of these two operators may be assumed to transform the other into itself multiplied by the commutator of order 2. The remaining pairs of independent generators may be assumed to be such that none of them generates the commutator subgroup but that each of the operators of a pair is the square of an operator of G and that two such operators are again non-commutative but are commutative with all of the other operators thus constructed.

We thus arrive at a G which has the given abelian group for its central and is not a direct product of two groups since its central is generated by the squares of its operators. When l is odd we may proceed similarly with the exception that the operator which does not appear in a pair may be assumed to be the only one of the set of independent generators which separately

---

* Received Aug. 3, 1937.

generates the commutator subgroup and is a non-square. This group is again not a direct product since if it were a direct product one of the factor groups would be abelian and would not involve the commutator of order 2. Hence there results the following theorem: *It is possible to construct a group of order 2^m which has an arbitrary abelian group whose order is of this form as its central and has a commutator subgroup of order 2 but is not the direct product of two groups.* Whenever two G's have centrals which are not simply isomorphic the G's are not simply isomorphic but there are also non-simply isomorphic G's which have simply isomorphic centrals.

The G's which have for their common central the group of order 2 are identical with the non-abelian groups which separately satisfy the condition that the squares of their operators constitute the group of order 2. It is known that there are three infinite systems of such groups.[1] We proceed to determine the G's which have for their centrals the group of type 1^n, where $n > 1$, and shall first impose the additional condition that each of the operators of order 2 contained in such a group is invariant. Let s_1 and s_2 represent two non-invariant operators of order 4. The group generated by s_1 and s_2 may be of order 8, 16, or 32. When it is of order 8 it is the quaternion group. When it is of order 16 it contains 12 operators of order 4 which have two distinct squares and hence it is completely determined. When it is of order 32 the squares of s_1 and s_2 do not generate the commutator subgroup and hence it is again completely determined.

Suppose now that G contains the quarternion group. Its subgroup composed of its operators which are commutative with all the operators of this quaternion group cannot involve any operator whose square is the commutator of order 2. If this subgroup involves two non-commutative operators of order 4 they therefore generate the given group of order 32. This is also the case when the subgroup of index 4 under this subgroup composed of all its operators which are commutative with each of these two non-commutative operators involves two non-commutative operators of order 4, etc. By continuing this process we finally arrive, when G is of finite order, at an abelian subgroup of type 1^n, n being of the form $2m + 1$, where $m + 1$ represents the number of these successive steps. Such a system can readily be constructed by starting with the abelian group of type 1^{2m+1} and inserting the quaternion group at an arbitrary stage of the process. The order of G is 2^{4m+3}, where m is an arbitrary positive integer or zero. This system of groups is characterized by the following properties. Each group has the group of order 2 for its commutator subgroup, involves the quaternion group, contains only operators of order 2 and 4 besides the identity, every operator of order 2 is in the central and every operator of order 4 is non-invariant, there is one and only one group for a given positive or zero value of m.

---

[1] G. A. Miller, *American Journal of Mathematics*, vol. 55, pp. 417-420.

When G does not involve two non-commutative operators of order 4 which generate the quaternion group but contains two such operators which generate the given group of order 16 then we can proceed similarly to construct an infinite system of groups which satisfy the conditions in question. In this case we arrive at the abelian group of type $1^{2(m+1)}$, and each of the successive groups, after the first, is again of order 32. This system is characterized just as the preceding one except that the quarternion group therein is replaced by the given group of order 16 and the order of G is $2^{4(m+1)}$. Finally, when every two non-invariant operators of G generate the given group of order 32 there results the third and last system of groups which are separately characterized by the facts that each group has the group of order 2 for its commutator subgroup, contains no operator whose order exceeds 4, all of its operators of order 2 are in its central while each of its operators of order 4 is non-invariant. The order of each of these groups is 2^{4m+5}, where m is a positive integer or zero, and there is one and only one such group for every such value of m. That is, *there are three and only three infinite systems of groups which separately satisfy the three conditions that each of them has the group of order* 2 *for its commutator subgroup, contains no operator whose order exceeds* 4, *all its operators of order* 2 *are in its central but none of its operators of order* 4 *appears therein.*

We proceed to consider the groups which are characterized by the conditions that their centrals are of type 1^n, $n > 1$, and that they separately involve non-invariant operators of order 2. Such a G is generated by its operators of order 4 since these operators could not generate a proper subgroup such that each of the remaining operators is of order 2 since this proper subgroup would be abelian and its operators would be transformed into their inverses by the remaining operators of G. When all the operators of order 2 contained in G are relatively commutative they generate an invariant abelian subgroup of G. As G is supposed to contain operators of order 2 which are not commutative with all of its operators of order 4 it results that it contains two non-commutative operators of order 4 which generate a group of order 16 involving non-invariant operators of order 2. Hence the following theorem: *If the central of a group whose commutator subgroup is of order* 2 *is of type* 1^n *and if this group involves non-invariant operators of order* 2 *but all of its operators of this order are relatively commutative then it contains invariantly* *the group of order* 16 *which is characterized by the fact that it contains exactly eight operators of order* 4 *which have two distinct squares.*

When the operators of order 2 contained in G generate an abelian subgroup then G can be constructed by successively extending this subgroup by operators of order 4 which are commutative with half of the operators of this subgroup and whose squares do not include the commutator of order 2 contained in G. This process of extending this subgroup and the resulting group is continued until we arrive at a group whose central is generated by the squares of the added operators of order 4. The extending operator in each case has a square which does not appear in the group of the squares of the operators of order 4 previously constructed. If the original abelian subgroup is of order 2^k the number of these extensions is $k - 1$ and the central of each extended group is half of the central of the preceding group. The smallest group in this system is the group of order 16 noted at the close of the preceding paragraph, and the central of each of these groups contains one and only one operator of order 2 which is not a square, namely the commutator of this order.

Suppose that a group whose commutator subgroup is of order 2 and whose central is of type 1^n is generated by its operators of order 2. Two non-commutative operators of order 2 contained in such a G generate the octic group. The subgroup of G composed of all its operators which are commutative with every operator of this octic group is also generated by its operators of order 2 since G is not a direct product. Hence there results the theorem that if a group whose commutator subgroup is of order 2 and whose central is of type 1^n is not a direct product it belongs to the infinite system of groups which is characterized by the facts that more than half of the operators of each group are of order 2 and that the squares of the operators of each group constitute the group of order 2. It therefore results that the groups under consideration are not separately generated by their operators of order 2.

The groups whose commutator subgroups are of order 2 and whose centrals are of type 1^n, $n > 1$, therefore have the property that they involve proper subgroups generated by their operators of order 2 whenever they involve non-invariant operators of this order. Such a proper subgroup involves invariant operators of order 2 besides the commutator of this order and the squares of its operators of order 4, and G can be constructed by extending such a subgroup by operators of order 4 which have different squares and whose squares do not include the commutator of order 2. Such an extending operator is non-commutative with an invariant operator of order 2 of this subgroup which is not the commutator nor a square. If the resulting group contains an invariant operator of order 2 which is not a commutator or a square this extension is repeated until the resulting group contains no invariant operator of order 2 besides the commutator of this order and those which are squares. By these methods all the groups whose centrals are of type 1^n, $n > 1$, and whose commutator subgroups are of order 2 can be constructed.

It was noted above that the squares of all the operators of G appear in the central of G. When the commutator of order 2 in G is the square of an invariant operator of G then the squares of the operators of G generate a subgroup composed of operators which are squares. This results from the fact that if s and t represent any two operators of such a G then s^1t^2 is the square of an operator of G since when s and t are non-commutative then s^2t^2 is the square of st multiplied by an invariant operator of order 4 which generates the commutator of order 2. The central of G may contain operators whose order exceeds 2 which are not squares but generate the commutator of order 2. It may also contain operators which are not squares and do not generate the commutator of order 2. In the latter case such an operator generates also the square of a non-invariant operator of G since G would otherwise be a direct product. Hence there results the following theorem: *If the central of a group whose commutator subgroup is of order* 2 *contains operators which are not squares of other operators of the group then such an operator either generates the commutator of order* 2 *or the square of a non-invariant operator whose order exceeds* 2.

If s is a non-invariant operator of lowest order in G then s can not generate the commutator of order 2 if G contains an operator of larger order which generates this commutator. If there is another operator in G of the same order as s which is non-commutative with s then this operator can not generate the commutator of order 2 in G if its order exceeds 4. Hence it results that we can select a set of operators of G in the following manner: The first two are non-commutative operators of lowest order in G, the second two are non-commutative operators of lowest order in the subgroup of index 4 composed of all the operators of G which are commutative with the first two selected operators, etc. Then if one of these operators generates the commutator of order 2 contained in G none of the other operators of the set can have this property unless the two operators are of order 4.

From what precedes it results that the groups whose commutator subgroup is of order 2 can be divided into two categories. In one of these the commutator of order 2 is generated only by operators in the central while in the other this commutator is generated by some non-invariant operator. In the latter case a set of independent generators of each of the groups can be so selected that one and only one of them generates the commutator of order 2 unless two of them appear in the quaternion group. In this case none of the remaining operators of the set generates the commutator of order 2. Hence the construction of these groups is to a large extent reduced to the construction of groups generated by two non-invariant operators in addition to the central. Therefore we proceed to consider this special case.

If these two non-invariant operators are in the quaternion group then the central may be any cyclic group which generates the commutator of order 2 but it can be no other group. When the order of this cyclic group exceeds 2 the group is also generated by two non-invariant operators which appear in the octic group together with the same central cyclic group. When each of the two non-invariant generating operators has an order which exceeds 4 and one of them generates the commutator of order 2 the central may be an arbitrary abelian group having at most two invariants, which are at least equal respectively to half the orders of these two independent generators. Finally, when neither of the two non-invariant generating operators generates the commutator of order 2 the central may be an arbitrary abelian group having at most three invariants. One of the three independent generators of the central, in the case that there are three such generators, then generates the commutator of order 2 and has an arbitrary order while the orders of the other two are respectively at least equal to one-half of the orders of the two non-invariant generators of *G*.

UNIVERSITY OF ILLINOIS.

CONVERGENCE OF A SEQUENCE OF LINEAR TRANSFORMATIONS.*

By M. H. Ingraham and M. C. Wolf.[1]

The purpose of the present paper is to study conditions on a sequence $\{A_i\}$ of $n \times n$ matrices with elements in the complex field, such that if an infinite sequence of linear transformations with matrices $\{A_i\}$ is applied to any bounded region of an n-dimensional vector space, the region converges uniformly to the origin. Certain phases of the problem are discussed in general for n-space and carried out in detail for $n = 2$.

The work on this paper was suggested by a problem in the theory of genetics.[2]

I. Sufficient conditions in the general case. An n-dimensional vector ξ after the application of a linear transformation with $n \times n$ matrix A is the vector $A\xi$ of length $\sqrt{\bar{\xi}'\bar{A}'A\xi}$ with direction components which are in general different from those of ξ. If $A\xi_i = \lambda_i\xi_i$, then ξ_i is an invariant direction (characteristic direction or vector) of the transformation with matrix A, and is said to correspond to the characteristic value λ_i. In general, for convenience, the vectors will be taken unitary.

---

* Presented to the American Mathematical Society, April 9, 1937. Received by the Editors, July 15, 1937.

[1] The work of M. C. Wolf was supported by a grant from the Wisconsin Alumni Research Foundation.

[2] For a purely Mendelian case of a characteristic determined by one gene, there are three possible types: 1) homozygous dominant, 2) heterozygous, and 3) homozygous recessive. The proportions of these three types in a population can be represented by a point in a plane, or in the case of n genes by a point in space of $2n$ dimensions.

Under certain hypothesis as to selective mating and productivity the coördinates of the point representing one generation can be given as quotients of quadratic forms in the coördinates of the point representing the preceding generation.

Invariant points represent conditions of equilibrium. These may be of stable or unstable type depending on whether or not the invariant point is the limit of the iterate of the transformation operating on all neighboring points. This may be studied by means of the iterates of the Jacobian of the transformation at the invariant point. As the transformations for any case but that of an "infinite population" vary slightly from generation to generation within very small limits, a better picture of the actual case is obtained by studying the products of a finite number of nearly equal transformations taken from an infinite sequence of such transformations rather than the iterates of a single transformation.

When $\bar{\xi}'_1\xi_1 = \bar{\xi}'_2\xi_2 = 1$, then $|\bar{\xi}'_1\xi_2| \leq 1$. Hence for any matrices R and S and any unitary vector ξ, it follows that for all unitary η,

$$|\bar{\xi}'RS\xi| \leq \sqrt{\max \bar{\eta}'\bar{R}'R\eta \quad \max \bar{\eta}'\bar{S}'S\eta}. \tag{1}$$

If A is an $n \times n$ matrix with characteristic values λ_i $(i = 1, 2, \cdots, n)$, and if the columns of a non-singular matrix T are a set of corresponding characteristic directions $\xi_1, \xi_2, \cdots, \xi_n$, then $A = TBT^{-1}$ where

$$B = \begin{pmatrix} \lambda_1 & & & & 0 \\ & \lambda_2 & & & \\ & & \cdot & & \\ & & & \cdot & \\ & & & & \cdot \\ 0 & & & & \lambda_n \end{pmatrix}.$$

If $\bar{\xi}'\xi = 1$, and if $|\lambda_1| \geq |\lambda_2| \geq \cdots \geq |\lambda_n|$, then $\bar{\xi}'\bar{B}'B\xi \leq N(\lambda_1)$, where $N(\lambda_1)$ is the norm of λ_1. Since $A = TBT^{-1}$ it follows from (1) that

$$\bar{\xi}'\bar{A}'A\xi \leq N(\lambda_1)K^2 \tag{2}$$

where K is a function of the columns of T, and hence depends only on the characteristic directions of A. K may be taken greater than unity. It is important to note that $N(\lambda_1)K^2$ is an upper bound for the Hermitian form $\bar{\xi}'\bar{R}'R\xi$ for unitary ξ associated with any matrix R for which the characteristic directions are given by the above matrix T and for which $N(\lambda_1)$ is the maximum norm of the characteristic values.

THEOREM 1. *The unitary vectors converge uniformly toward the origin under a sequence of transformations with matrices $\{A_i\}$ where $A_i = A + E_i$, if $\{E_i\}$ is a sequence of $n \times n$ matrices such that $\bar{\xi}'\bar{E}'_iE_i\xi \leq H^2$ for all i and all unitary ξ, $N(\lambda_1)$ and K are the constants described above, and if H is positive and satisfies the condition $[N(\lambda_i) + KH] < 1$.*

Let

$$C_k = (A + E_k)(A + E_{k-1}) \cdots (A + E_2)(A + E_1) = \sum_{i=0}^{k} P_i$$

where P_i is a sum of $\binom{k}{i}$ terms; each term of P_i is a product of k matrices in which the matrix A is a factor $k - i$ times and the remaining i factors are i of the matrices $E_1, E_2, \cdots, E_k$. For example, if $k = 9$ and $i = 3$, $A^2E_7AE_5E_4A^3$ is a typical term of P_3.

$$\bar{C}'_k = \sum_{i=0}^{k} \bar{P}'_i$$

$$\bar{\xi}'\bar{C}'_kC_k\xi = \bar{\xi}' \sum_{l=0}^{2k} Q_l\xi = \sum_{l=0}^{2k} \bar{\xi}'Q_l\xi$$

where each Q_l is a sum of $\binom{2k}{l}$ products of the form $X_1X_2\cdots X_kY_1Y_2\cdots Y_k$. Of the X_i, $k-s$ are $\bar{A}'$ and s are certain of the $\bar{E}'$; $k-t$ of the Y_i are A and the remaining t are certain of the E_i; $t+s=l$. To determine a bound for $|\bar{\xi}'Q_l\xi|$ apply (1). Let $N(\lambda_1)=M^2$. The bound for $\sqrt{\bar{\xi}'\bar{A}'^iA^i\xi}$ as determined from (2) is $\sqrt{[N(\lambda_1)]^iK^2}=M^iK$; $\sqrt{\bar{\xi}'\bar{E}'_iE_i\xi}\leqq H$; K enters the bound for $\bar{\xi}'Q_l\xi$ for every separate power A^p or $\bar{A}'^p$ in Q_l, but the factors $\bar{A}'^p$ in the product $X_1X_2\cdots X_k$ are separated by $\bar{E}'_i$'s at most $s+1$ times and the factors A^p in $Y_1Y_2\cdots Y_k$ are separated by E_i's at most $t+1$ times, hence it follows that if $K\geqq 1$

$$|\bar{\xi}'Q_l\xi|<\binom{2k}{l}M^{k-s}K^{s+1}H^s\cdot M^{k-t}K^{t+1}H^t$$

$$|\bar{\xi}'Q_l\xi|<K^2\cdot\binom{2k}{l}M^{2k-l}H^lK^l.$$

For example, suppose a term of Q_5, for $k=8$, is

$$T_5=\bar{A}'^3\bar{E}'_4\bar{E}'_5\bar{A}'\bar{E}'_7\bar{A}'\cdot A^4E_4A^2E_1.$$

$$|\bar{\xi}'T_5\xi|\leqq(M^3K\cdot H\cdot H\cdot MK\cdot H\cdot MK)\cdot(M^4K\cdot H\cdot M^2K\cdot H)=M^{11}K^5H^5$$

$$|\bar{\xi}'T_5\xi|<K^2\cdot M^{11}K^5H^5.$$

Clearly when $K\geqq 1$

$$\bar{\xi}'\bar{C}'_kC_k\xi<K^2(M+HK)^{2k}.$$

If $K<1$, then

$$\bar{\xi}'\bar{C}'_kC_k\xi<(M+HK)^{2k}.$$

If $(M+HK)<1$, $\lim_{k\to\infty}(M+HK)^{2k}\to 0$. Hence if $|\lambda_i|<1$ and if $A+E_i$ differs but slightly from A, that is, if $\bar{\xi}'\bar{E}'_iE_i\xi\leqq H^2$ is sufficiently small, the region about the origin will converge uniformly to the origin under the sequence of transformations with matrices $\{A_i\}$.

In the remainder of this section more detailed conditions for convergence and divergence of vector spaces will be given.

THEOREM 2. *If $\{g_i(A)\}$ is a sequence of polynomials in the above restricted matrix A, and if $\lim_{m\to\infty}N(\rho_m)\to 0$ where $N(\rho_m)$ is the maximum norm of the characteristic values of $\prod_{i=1}^{m}g_i(A)$, then any bounded portion of space* *tends uniformly toward the origin under the sequence of transformations with matrices* $\{g_i(A)\}$.

This follows from (2) since the characteristic directions of $\prod_{i=1}^{m} g_i(A)$ are those of A; furthermore, if λ_i is a characteristic value of A, $g_j(\lambda_i)$ is a characteristic value of $g_j(A)$ and $\prod_{j=1}^{m} g_j(\lambda_i)$ is a characteristic value of $\prod_{j=1}^{m} g_j(A)$.

COROLLARY 2. 1. *If A, restricted as above, is a matrix with characteristic values λ_i such that $N(\lambda_i) < 1$, then under iteration of the transformation whose matrix is A, the whole space tends uniformly toward the origin, and if $N(\lambda_i) > 1$, the space diverges.*

THEOREM 3. *If $\{A_i\}$ is a sequence of matrices with characteristic values λ_{ij}, the condition that $|\lambda_{ij}| < 1$, for every i and j, is not sufficient to insure that a bounded region in space will tend to the origin under the transformation with matrices $\{A_i\}$.*

An example of such a sequence may be constructed by iteration of two transformations alternately. The first takes a unitary vector ξ into η where $\bar{\eta}'\eta = k^2 > 1$ and the second rotates η into $l\xi$ where $k^2 > l^2 > 1$.

On the other hand, there exist sequences $\{A_i\}$ of matrices for which the norm of one characteristic value for every A_i is greater than unity, yet the unitary vectors converge uniformly to the origin. The sequence which is alternately A_1 and A_2 is an example of such when

$$A_1 = \begin{pmatrix} 1/3 & 0 \\ 0 & 3/2 \end{pmatrix} \qquad A_2 = \begin{pmatrix} 3/2 & 0 \\ 0 & 1/3 \end{pmatrix}.$$

THEOREM 4. *If $\{A_i\}$ is a sequence of matrices such that*

$$\|A_i\| = |\lambda_{i1}\lambda_{i2}\cdots\lambda_{in}| > 1,$$

then the set of unitary vectors will not converge to the origin under the sequence of transformations with matrices $\{A_i\}$.

Let $B_m = A_m A_{m-1}\cdots A_1$.

$$\|B_m\| = \|A_m\|\cdot\|A_{m-1}\|\cdots\|A_1\| > 1.$$

It suffices to prove that there exists at least one vector η such that $\bar{\eta}'\bar{B}'_m B_m \eta > 1$ for infinitely many values of m. Let

$$C_m = \begin{bmatrix} \sigma_1(m) & & & & 0 \\ & \sigma_2(m) & & & \\ & & \cdot & & \\ & & & \cdot & \\ 0 & & & & \sigma_n(m) \end{bmatrix} = \bar{T}'_m \bar{B}'_m B_m T_m,$$

$\bar{T}'_m T_m = I$. The set of functional values of $\bar{\zeta}' C_m \zeta$ is equal to those of $\bar{\zeta}' \bar{B}'_m B_m \zeta$ for unitary ζ. Consider real unitary vectors η with coördinates $x_1, x_2, \cdots, x_n$.

Suppose $x_1 \neq 0$ and

$$(3) \qquad \begin{gathered} \sigma_1(m) \geq \sigma_2(m) \geq \cdots \geq \sigma_n(m) \geq 0. \\ | C_m | = \sigma_1(m)\sigma_2(m) \cdots \sigma_n(m) > 1 \end{gathered}$$

is an increasing function of m. Hence $| C_m | > k > 1$ for some k and $\sigma_1(m) > \sqrt[n]{k} > 1$. From (3), $\bar{\eta}' C_m \eta \geq \sigma_1(m) x_1^2$. If $x_1 > \frac{1}{\sqrt[2n]{k}}$, $\bar{\eta}' C_m \eta > 1$. It follows that there exists a vector ζ_m and an ε_k, where ε_k is independent of m, such that $\bar{\zeta}'_m \bar{B}'_m B_m \zeta_m > 1$ and $\bar{\zeta}'_i \bar{B}'_m B_m \zeta_i > 1$ for all ζ_i for which

$$\left| \frac{\bar{\zeta}'_i \zeta_m}{\sqrt{\bar{\zeta}'_i \zeta_i \cdot \bar{\zeta}'_m \zeta_m}} - 1 \right| < \varepsilon_k.$$

The infinite sequence of vectors ζ_m $(m = 1, 2, \cdots)$ has a limit vector ζ. Let ζ_{1i} be an infinite subsequence of ζ_m such that

$$\left| \frac{\bar{\zeta}' \zeta_{1i}}{\sqrt{\bar{\zeta}' \zeta \cdot \bar{\zeta}'_{1i} \zeta_{1i}}} - 1 \right| < \varepsilon_k,$$

then $\bar{\zeta}' \bar{B}'_{1i} B_{1i} \zeta > 1$ for an infinite subsequence $\{B_{1i}\}$ of $\{B_i\}$, as was to be proved.

The proof of the next theorem is similar to that of Theorem 2.

THEOREM 5. *If* $\{A_i\}$ *is a sequence of matrices such that the characteristic values* λ_{ij} $(j = 1, 2, \cdots, n)$ *for all* i *correspond respectively to the linearly independent invariant directions* $\xi_1, \xi_2, \cdots, \xi_n$ *and if* $| \lambda_{ij} | < k < 1$, *the unitary vectors converge uniformly to the origin.*

The point of view of this paper is intimately connected with the Cauchy-Riemann conditions of analytic function theory. A 2×2 matrix $A = (a_{ij})$ may be interpreted as the Jacobian matrix of a transformation $U = u(x, y)$ and $V = v(x, y)$. At a point the Cauchy-Riemann conditions are that

$a_{11} = a_{22}$ and $a_{12} = -a_{21}$. If $\begin{pmatrix} 1 \\ m_1 \end{pmatrix}$ and $\begin{pmatrix} 1 \\ m_2 \end{pmatrix}$ are the invariant directions of A, m_1 and m_2 satisfy the equation $x^2 + 1 = 0$ unless $a_{12} = 0$, in which case A is of the form λI where λ is real. The Cauchy-Riemann conditions are therefore equivalent to having

$$\xi_1 = \begin{pmatrix} 1 \\ i \end{pmatrix} \quad \text{and} \quad \xi_2 = \begin{pmatrix} 1 \\ -i \end{pmatrix}$$

as characteristic vectors and λ_1 and $\lambda_2 = \bar{\lambda}_1$ as characteristic values. It is clear that these conditions give conformal representation, since if

$$\eta_1 = c_{11}\xi_1 + c_{12}\xi_2 \quad \text{and} \quad \eta_2 = c_{21}\xi_1 + c_{22}\xi_2,$$

then

$$\frac{\bar{\eta}'\bar{A}'A\eta_2}{\sqrt{\bar{\eta}'_1\bar{A}'A\eta_1}\,\sqrt{\bar{\eta}'_2\bar{A}'A\eta_2}} = \frac{\bar{\eta}'_1\eta_2}{\sqrt{\bar{\eta}'_1\eta_1}\,\sqrt{\bar{\eta}'_2\eta_2}}$$

which is the condition that the angle between η_1 and η_2 is preserved (except possibly for sign). This equality follows since $\bar{\xi}'_1\xi_2 = \bar{\xi}'_2\xi_1 = 0$, whence also for $i, j = 1, 2$

$$\bar{\eta}'_i\bar{A}'A\eta_j = \lambda_1\bar{\lambda}_1\bar{\eta}'_i\eta_j.$$

II. Characteristic directions of products in real 2-space. It is seen that the conditions under which the whole space under successive transformations tends toward the origin depend not only on the maximum stretch under any one transformation, but also upon the relations of the characteristic directions of the various transformations. It is plausible therefore that if the characteristic vectors of a series of matrices were nearly the same, results could be secured which are analogous to the case in which they are equal. Therefore this section is devoted to the study of the characteristic vectors and values of the product of 2 × 2 matrices.

Suppose A and B are 2 × 2 matrices with characteristic values λ_1 and λ_2, and γ_1 and γ_2 respectively where $|\lambda_1| \geq |\lambda_2|$ and $|\gamma_1| \geq |\gamma_2|$. Suppose ξ_1 and ξ_2 are the characteristic directions of A corresponding to λ_1 and λ_2 respectively, and $\xi_1 + \delta_1\xi_2$ and $\xi_2 + \delta_2\xi_1$ are the characteristic directions of B corresponding to γ_1 and γ_2 respectively, where $\bar{\xi}'_1\xi_1 = \bar{\xi}'_2\xi_2 = 1$ and ξ_1 and ξ_2 are linearly independent.

$$B(\xi_1 + \delta_1\xi_2) = \gamma_1(\xi_1 + \delta_1\xi_2).$$

Hence

$$B\xi_1 = \gamma_1\xi_1 + \delta_1\gamma_1\xi_2 - \delta_1 B\xi_2.$$

Similarly

$$B\xi_2 = \gamma_2\xi_2 + \delta_2\gamma_2\xi_1 - \delta_2 B\xi_1.$$

Therefore

$$B\xi_1 = (\gamma_1 - \delta_1\delta_2\gamma_2)\xi_1 + \delta_1(\gamma_1 - \gamma_2)\xi_2 + \delta_1\delta_2 B\xi_1.$$

If $\delta_1\delta_2 \neq 1$, then

$$B\xi_1 = \left(\frac{1}{1-\delta_1\delta_2}\right)\{(\gamma_1 - \delta_1\delta_2\gamma_2)\xi_1 + \delta_1(\gamma_1 - \gamma_2)\xi_2\}.$$

Also

$$B\xi_2 = \left(\frac{1}{1-\delta_1\delta_2}\right)\{(\gamma_2 - \delta_1\delta_2\gamma_1)\xi_2 + \delta_2(\gamma_2 - \gamma_1)\xi_1\}.$$

Suppose a characteristic vector of the product BA is of the form $\xi_1 + r\xi_2$, r is then a function of λ_i, γ_i, δ_i $(i = 1, 2)$. Suppose σ is the characteristic value of BA corresponding to $\xi_1 + r\xi_2$. Then

$$BA(\xi_1 + r\xi_2) = \sigma(\xi_1 + r\xi_2). \tag{4}$$

But

$$\begin{aligned} BA(\xi_1 + r\xi_2) &= B(\lambda_1\xi_1 + r\lambda_2\xi_2) \\ &= \left(\frac{1}{1-\delta_1\delta_2}\right)\{[\lambda_1(\gamma_1 - \delta_1\delta_2\gamma_2) + r\lambda_2\delta_2(\gamma_2 - \gamma_1)]\xi_1 \\ &\quad + [\lambda_1\delta_1(\gamma_1 - \gamma_2) + r\lambda_2(\gamma_2 - \delta_1\delta_2\gamma_1]\xi_2\}. \end{aligned} \tag{5}$$

Since ξ_1 and ξ_2 are linearly independent, it follows from (4) and (5) that

$$r = \frac{\lambda_1\delta_1(\gamma_1 - \gamma_2) + r\lambda_2(\gamma_2 - \delta_1\delta_2\gamma_1)}{\lambda_1(\gamma_1 - \delta_1\delta_2\gamma_2) + r\lambda_2\delta_2(\gamma_2 - \gamma_1)}.$$

Hence

$$\begin{aligned} f_1(r) \equiv \lambda_2\delta_2(\gamma_2 - \gamma_1)r^2 + \{\lambda_1(\gamma_1 - \delta_1\delta_2\gamma_2) - \lambda_2(\gamma_2 - \delta_1\delta_2\gamma_1)\}r \\ + \lambda_1\delta_1(\gamma_2 - \gamma_1) = 0. \end{aligned} \tag{6}$$

Similarly if $\xi_2 + t\xi_1$ is a characteristic vector for BA, then

$$\begin{aligned} f_2(t) \equiv \lambda_1\delta_1(\gamma_1 - \gamma_2)t^2 + \{\lambda_2(\gamma_2 - \delta_1\delta_2\gamma_1) - \lambda_1(\gamma_1 - \delta_1\delta_2\gamma_2)\}t \\ + \lambda_2\delta_2(\gamma_1 - \gamma_2) = 0. \end{aligned} \tag{7}$$

If only those matrices are studied for which λ_i, γ_i, ξ_i, δ_i are real, $|\delta_i| < 1$, $|\lambda_2/\lambda_1| < k < 1$ and $|\gamma_2/\gamma_1| < k < 1$, then bounds for the characteristic directions of the product BA depend upon the relative signs of the determinants $|B| = \gamma_1\gamma_2$ and $|A| = \lambda_1\lambda_2$.

Case I. Suppose $\lambda_1\lambda_2 \geqq 0$ and $\gamma_1\gamma_2 \geqq 0$. Since $f_1(0) = \lambda_1\delta_1(\gamma_2 - \gamma_1)$ and $f_1(\delta_1) = \delta_1\gamma_2(1 - \delta_1\delta_2)(\lambda_1 - \lambda_2)$, the sign of $f_1(0)$ is the sign of $-\lambda_1\gamma_1\delta_1$ and that of $f_1(\delta_1)$ is the same as $\delta_1\lambda_1\gamma_2$. Since λ_1 and λ_2 are of the same sign, and γ_1 and γ_2 are of the same sign, there is a root of (6) between 0 and δ_1. Similarly, there is a root of (7) between 0 and δ_2.

Case II. If, however, $\lambda_1\lambda_2 \geqq 0$ and $\gamma_1\gamma_2 \leqq 0$, there exists a root of (7) between 0 and δ_2 but $f_1(0)$ and $f_1(\delta_1)$ are of the same sign. Consider

$$f_1(2\delta_1) = \delta_1[\lambda_1\gamma_1 + 2\lambda_2(-\gamma_2) - \lambda_1(-\gamma_2)] + \lambda_1\gamma_1K_1$$
$$|K_1| \leqq 4(k^2 + k)\,|\delta_1{}^2\delta_2|.$$

The term $[\lambda_1\gamma_1 + 2\lambda_2(-\gamma_2) - \lambda_1(-\gamma_2)]$ bears the sign of $\lambda_1\gamma_1$. Since in this case the sign of $\dfrac{f_1(0)}{\lambda_1\gamma_1}$ is the sign of $-\delta_1$, there is a root of (6) between 0 and $2\delta_1$ if the sign of $\dfrac{f_1(2\delta_1)}{\gamma_1\lambda_1}$ has the sign of δ_1. This is the case if the values of δ_1 and δ_2 are sufficiently small, so that K_1 will not affect the sign of $\dfrac{f_1(2\delta)}{\gamma_1\lambda_1}$. A bound for these values of δ_1 and δ_2 may be found depending upon k alone.

Case III. If $\lambda_1\lambda_2 \leqq 0$ and $\gamma_1\gamma_2 \geqq 0$, there exists a root of (6) between 0 and δ_1, but $f_2(0)$ and $f_2(\delta_2)$ are of the same sign.

$$f_2(-2\delta_2) = \delta_2[3\gamma_2(-\lambda_2) + 2\lambda_1\gamma_1 - \gamma_1(-\lambda_2)] + \lambda_1\gamma_1K_2$$
$$|K_2| \leqq 4(1 + 2k)\,|\delta_1\delta_2{}^2|.$$

The sign of $\dfrac{f_2(-2\delta_2)}{\lambda_1\gamma_1}$ is that of δ_2 if δ_1 and δ_2 are small enough so that K_2 does not affect the sign of $\dfrac{f_2(-2\delta_2)}{\lambda_1\gamma_1}$. Then there is a root of (7) between 0 and $-2\delta_2$ because the sign of $\dfrac{f_2(0)}{\lambda_1\gamma_1}$ is the sign of $-\delta_2$.

Case IV. If $\lambda_1\lambda_2 \leqq 0$ and $\gamma_1\gamma_2 \leqq 0$, consider

$$f_1\left(\frac{\delta_1}{1-k}\right) = \frac{\lambda_1\gamma_1\delta_1}{1-k}\left[1 - \frac{\lambda_2\gamma_2}{\lambda_1\gamma_1} + (1-k)\frac{\gamma_2}{\gamma_1} - (1-k)\right] + \gamma_1\lambda_1K_3$$
$$|K_3| \leqq \frac{3k - k^2}{(1-k)^2}\,|\delta_1{}^2\delta_2|.$$

The sign of $\dfrac{f_1\left(\dfrac{\delta_1}{1-k}\right)}{\lambda_1\gamma_1}$ is that of δ_1 if δ_1 and δ_2 are small enough so that K_3 does not affect the sign of $f_1\left(\dfrac{\delta_1}{1-k}\right)$. The sign of $\dfrac{f_1(0)}{\lambda_1\gamma_1}$ is that of $-\delta_1$, hence there is a root of (6) between 0 and $\dfrac{\delta_1}{1-k}$. The sign of $\dfrac{f_2(0)}{\lambda_1\gamma_1}$ is that of $-\delta_2$. Consider

$$f_2\left(\frac{-k\delta_2}{1-k}\right)=\frac{\lambda_1\gamma_1\delta_2}{1-k}\left\{k+\frac{\lambda_2}{\lambda_1}-k\frac{\lambda_2}{\lambda_1}-\frac{\lambda_2\gamma_2}{\lambda_1\gamma_1}\right\}+\lambda_1\gamma_1K_4$$

$$|K_4|\leq\frac{k^2(3-k)}{(1-k)^2}\,|\delta_1\delta_2^2|.$$

Neglecting K_4, the sign of $\dfrac{f_2\left(\dfrac{-k\delta_2}{1-k}\right)}{\lambda_1\gamma_1}$ is that of δ_2, hence there is a root of (7) between 0 and $\dfrac{-k\delta_2}{1-k}$ when δ_1 and δ_2 are sufficiently small, a bound for them depending upon the value of k alone.

Since the smaller roots of (6) and (7) are such that in Cases I, II, III, $|r|<|2\delta_1|$ and $|t|<|2\delta_2|$, and in Case IV $|r|<\dfrac{|\delta_1|}{1-k}$ and $|t|<\dfrac{k\,|\delta_2|}{1-k}$, when δ_1 and δ_2 are small quantities, the quadratic terms in r and t of (6) and (7) may be neglected as well as quadratic terms in δ_1 and δ_2. Consequently r is of the order of $\dfrac{\lambda_1\delta_1(\gamma_1-\gamma_2)}{\lambda_1\gamma_1-\lambda_2\gamma_2}$ and t is of the order of $\dfrac{\lambda_2\delta_2(\gamma_2-\gamma_1)}{\lambda_2\gamma_2-\lambda_1\gamma_1}$.

Equating coefficients in equations (4) and (5), the characteristic value corresponding to $\xi_1+r\xi_2$ becomes

$$\sigma=\frac{\lambda_1\gamma_1}{1-\delta_1\delta_2}\left\{1-\delta_1\delta_2\frac{\gamma_2}{\gamma_1}+r\delta_2\frac{\lambda_2}{\lambda_1}\left(\frac{\gamma_2}{\gamma_1}-1\right)\right\}. \tag{8}$$

Similarly the characteristic value ρ corresponding to $\xi_2+t\xi_1$ is

$$\rho=\frac{\lambda_2\gamma_2}{1-\delta_1\delta_2}\left\{1-\delta_1\delta_2\frac{\gamma_1}{\gamma_2}+t\delta_1\frac{\lambda_1}{\lambda_2}\left(\frac{\gamma_1}{\gamma_2}-1\right)\right\}. \tag{9}$$

$$\sigma=\frac{\lambda_1\gamma_1}{1-\delta_1\delta_2}\{1-K_5\}$$

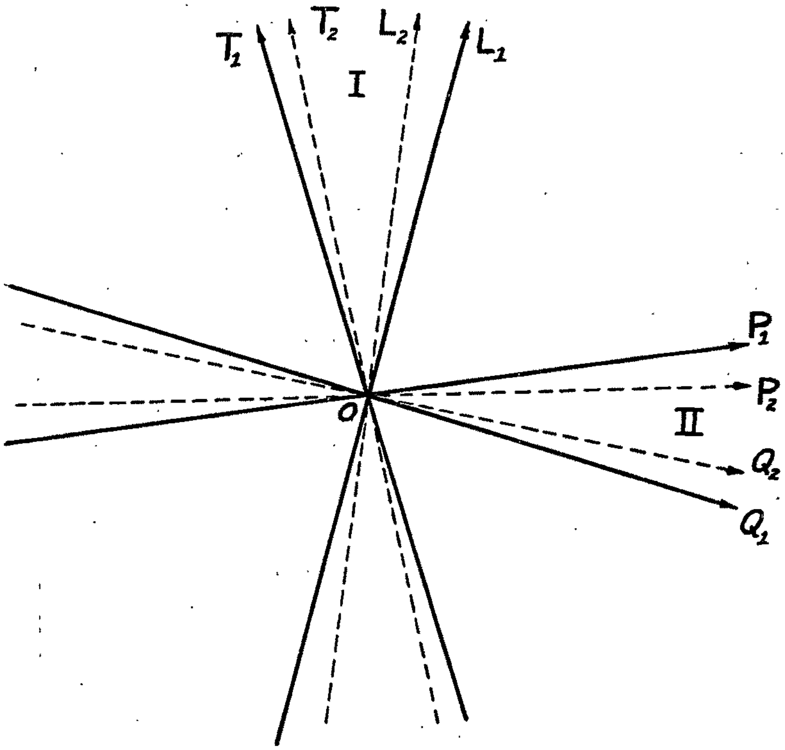
$$|K_5|<|\delta_1\delta_2|\,k+|r|\;|\delta_2|\,k(1-k).$$

Hence for small δ_1 and δ_2, σ is of the order of $\lambda_1\gamma_1$.

$$\frac{\rho}{\sigma}=\frac{\dfrac{\lambda_2\gamma_2}{\lambda_1\gamma_1}-\delta_1\delta_2\dfrac{\lambda_2}{\lambda_1}+t\delta_1\left(1-\dfrac{\gamma_2}{\gamma_1}\right)}{1-\delta_1\delta_2\dfrac{\gamma_2}{\gamma_1}+r\delta_2\dfrac{\lambda_2}{\lambda_1}\left(\dfrac{\gamma_2}{\gamma_1}-1\right)}$$

$$=\frac{\dfrac{\lambda_2\gamma_2}{\lambda_1\gamma_1}-K_6}{1-K_5}$$

$$|K_6|<|\delta_1\delta_2|k+(1+k)|t|\,|\delta_1|.$$



Hence $\dfrac{\rho}{\sigma}$ is of the order of $\dfrac{\lambda_2\gamma_2}{\lambda_1\gamma_1}$ for small δ_1 and δ_2 from which it follows that ρ is of the order of $\lambda_2\gamma_2$.

The nature of the results in Cases I-IV were arrived at geometrically. The simplest case is illustrated by the accompanying figure. Let OT_1 and OP_1 be the directions ξ_1 and ξ_2 corresponding to the larger and smaller roots of A, λ_1 and λ_2 respectively. Let OL_1 and OQ_1 be the directions $\xi_1+\delta_1\xi_2$ and

$\xi_2 + \delta_2\xi_1$, corresponding to the larger and smaller roots of B, γ_1 and γ_2 respectively. In this case $\lambda_1 > \lambda_2 > 0$ and $\gamma_1 > \gamma_2 > 0$. Under the transformation with matrix A, OT_1 and OP_1 are invariant but OL_1 and OQ_1 move into regions I and II, along the directions OL_2 and OQ_2. Under the transformation with matrix B, OT_1 and OP_1 move into regions I and II along the directions OT_2 and OP_2. From continuity considerations it follows that invariant directions for the product BA lie in regions I and II. In Cases II-III and less simple cases of I geometric arguments were also helpful but the diagrams were complicated by reflection caused by the occurrence of negative λ_i and γ_i.

THEOREM 6. *If $\{A_i\}$ is a sequence of second order matrices with the characteristic values of A_i equal to λ_{i1} and λ_{i2}, where $\left|\frac{\lambda_{i2}}{\lambda_{i1}}\right| < k < 1$, if ξ_1 and ξ_2 are two directions in space, then for every arbitrarily chosen number $\varepsilon > 0$ there exist numbers $\delta_{\epsilon k} > 0$ and $\delta'_{\epsilon k} > 0$ such that if the characteristic directions corresponding to λ_{i1} and λ_{i2} lie in the ranges $\xi_1 \pm \delta_{\epsilon k}\xi_2$ and $\xi_2 \pm \delta'_{\epsilon k}\xi_1$ respectively, then the characteristic directions of any product of a finite sequence of these matrices lie in the ranges $\xi_1 \pm \epsilon\xi_2$ and $\xi_2 \pm \epsilon\xi_1$. It is assumed that all quantities under consideration are real.*

Suppose it is known that for each of a sequence of matrices the characteristic directions corresponding to λ_{i1} and λ_{i2} lie within the ranges $\xi_1 \pm \delta_1\xi_2$ and $\xi_2 \pm \delta_2\xi_1$ respectively. Let $\delta_1 > 0$ and $\delta_2 > 0$. The product Q_m of m matrices may be expressed in the form

$$Q_m = N_m \cdot P_{k+1} \cdot M_k \cdot \cdot \cdot M_2 \cdot P_2 \cdot M_1 \cdot P_1,$$

where P_j is a product of matrices A_i each with positive determinant; M_j is a product of matrices A_i with the determinants of the first and last matrices negative, and those determinants of the intervening matrices positive. Let $M_j = A_{j1}T_jA_{j2}$. The determinants of T_j and M_j are positive. N_m is either the identity or else is a product of matrices A_i with only the right-hand factor having a negative determinant, all others being positive. Any P_j or T_j may be equal to the identity.

From Case I, for every P_j and T_j the characteristic directions corresponding to the larger and smaller characteristic values lie in the ranges $\xi_1 \pm \delta_1\xi_2$ and $\xi_2 \pm \delta_2\xi_1$ respectively. When δ_1 and δ_2 are properly restricted, the characteristic values of the product of t of the matrices A_i are approximately $\prod_{i=1}^{t} \lambda_{i1}$ and $\prod_{i=1}^{t} \lambda_{i2}$. Then $\left|\frac{\prod_{i=1}^{t}\lambda_{i1}}{\prod_{i=1}^{t}\lambda_{i2}}\right| < k < 1$. It follows from Case III that the characteristic direction corresponding to the larger characteristic value of T_jA_{j2} lies in $\xi_1 \pm \delta_1\xi_2$. Using the above theory of Cases I-IV, to compute bounds for the invariant direction corresponding to the smaller characteristic value it is only necessary to establish limits when the characteristic directions of T_j and A_{j2} are on the boundaries of the regions, $\xi_1 \pm \delta_1\xi_2$ and $\xi_2 \pm \delta_2\xi_1$. For example, suppose that the invariant directions of A_{j2} are $\xi_1 - \delta_1\xi_2$ and $\xi_2 + \delta_2\xi_1$ and the invariant directions of T_j are $\xi_1 + \delta_1\xi_2$ and $\xi_2 - \delta_2\xi_1$. To apply the results of Case III note that $\xi_1 + \delta_1\xi_2$ and $\xi_2 - \delta_2\xi_1$ lie in the directions

$$(\xi_1 - \delta_1\xi_2) + \frac{2\delta_1}{1 - \delta_1\delta_2}(\xi_2 + \delta_2\xi_1)$$

and

$$(\xi_2 + \delta_2\xi_1) - \frac{2\delta_2}{1 - \delta_1\delta_2}(\xi_1 - \delta_1\xi_2)$$

respectively. Then the invariant direction of the product corresponding to the smaller characteristic value lies between $\xi_2 + \delta_2\xi_1$ and

$$(\xi_2 + \delta_2\xi_1) + \frac{4\delta_2}{1 - \delta_1\delta_2}(\xi_1 - \delta_1\xi_2),$$

that is between $\xi_2 + \delta_2\xi_1$ and $\xi_2 + \dfrac{5 - \delta_1\delta_2}{1 - 5\delta_1\delta_2}\delta_2\xi_1$. Similarly, if the other possible bounds for the invariant directions of A_{j2} and T_j are considered, it follows that the direction of the product T_jA_{j2} is bounded by $\xi_2 \pm \mu_2\delta_2\xi_1$ where $\mu_2 = \left| \dfrac{5 - \delta_1\delta_2}{1 - 5\delta_1\delta_2} \right|$; μ_2 is of the order of 5 if δ_1 and δ_2 are small.

Bounds for the characteristic directions of $M_j = A_{j1}T_jA_{j2}$ are given through Case IV. For example, suppose the invariant direction corresponding to the larger root of T_jA_{j2} is $\xi_1 - \delta_1\xi_2$, and that of A_{j1} is $\xi_1 - \delta_1\xi_2$. Suppose the invariant direction corresponding to the smaller root of T_jA_{j2} is either of $\xi_2 \pm \mu_2\delta_2\xi_1$. To apply the results of Case IV, the direction $\xi_1 - \delta_1\xi_2$ must be expressed in the form

$$\xi_1 + \delta_1\xi_2 - \frac{2\delta_1}{1 \pm \mu_2\delta_1\delta_2}(\xi_2 \pm \mu_2\delta_2\xi_1).$$

It follows that the characteristic direction corresponding to the larger characteristic value of $A_{j1}T_jA_{j2}$ lies between $\xi_1 + \delta_1\xi_2$ and $\xi_1 - \tau_1\delta_1\xi_2$ where τ_1 is either

$$\frac{(1 + k) - (1 - k)\delta_1\delta_2\mu_2}{(1 - k) - (1 + k)\delta_1\delta_2\mu_2}$$

or

$$\frac{(1 + k) + (1 - k)\delta_1\delta_2\mu_2}{(1 - k) + (1 + k)\delta_1\delta_2\mu_2}.$$

If the other bounds for the characteristic directions of A_{j1} and T_jA_{j2} are considered, bounds for the characteristic directions of M_j are similarly found to be $\xi_1 \pm \tau_1\delta_1\xi_2$ and $\xi_2 \pm \tau_2\delta_2\xi_1$. The orders of τ_1 and τ_2 are less than $\frac{6}{1-k}$. Since the determinant of M_j is positive, the characteristic directions of M_jP_j lie within the larger of the two ranges given for M_j and P_j and since these ranges are independent of j, the characteristic directions of any number of products of M_jP_j lie in the same range. It is assumed that δ_1 and δ_2 are sufficiently small so that the approximation formulas hold to within the limits necessary for the above reasoning, and so that the characteristic values of a product of matrices closely approximate the product of the characteristic values. If an odd number of matrices with negative determinants is included in the product Q_m, that is, if the determinant of N_m is negative, the characteristic direction corresponding to the larger characteristic value of Q_m may fall outside of $\xi_1 \pm \tau_1\delta_1\xi_2$, but since the characteristic directions of N_m according to Case III are in the ranges $\xi_1 \pm \delta_1\xi_2$ and $\xi_2 \pm \mu_2\delta_2\xi_1$, it can be shown from Case II that the characteristic directions of Q_m lie in $\xi_1 \pm \sigma_1\delta_1\xi_2$ and $\xi_2 \pm \tau_2\delta_2\xi_1$. The bound $\xi_1 \pm \sigma_1\delta_1\xi_2$ is of the order of $\xi_1 + \frac{3-k}{1-k}\delta_1\xi_2$.

Hence if ε is an arbitrary number, sufficient conditions for the proof of Theorem 6 are that $\delta_{\epsilon k}$ and $\delta'_{\epsilon k}$ are sufficiently small to satisfy the conditions of Cases I-IV which depend upon k alone and also $\delta_{\epsilon k} < \frac{\varepsilon}{\sigma_1}$ and $\delta'_{\epsilon k} < \frac{\varepsilon}{\tau_2}$.

UNIVERSITY OF WISCONSIN.

THE INTERRELATIONS OF THE FUNDAMENTAL SOLUTIONS OF THE HYPERGEOMETRIC EQUATION; LOGARITHMIC CASE.*

By LYLE E. MEHLENBACHER.

1. The problem of this paper is to study the exact nature of the linear relations existing between the fundamental solutions of the Hypergeometric Differential Equation

$$(1) \qquad z(1-z)\frac{d^2y}{dz^2}+[\gamma-(\alpha+\beta+1)z]\frac{dy}{dz}-\alpha\beta y=0$$

in which we may consider the constants α, β, γ and the variable z as real or complex. The relations of this kind which exist in case each of the three regular singular points $z=0$, $z=1$, $z=\infty$ is non-logarithmic in character have already been considered by Forsyth,[1] Lindelöf[2] and Barnes.[3] What is proposed here is to determine the linear relations between the fundamental solutions of (1) when one or more of these solutions becomes logarithmic in character. We shall, for the sake of brevity, actually determine only those linear relations existing between the fundamental solutions about the point $z=0$ and the solutions about the point $z=\infty$ when one of the solutions about the latter point is logarithmic. The methods employed are readily applied to the cases in which either the point $z=0$, or the point $z=1$ is logarithmic. The complete set of interrelations in both the non-logarithmic cases and the logarithmic cases are included in the author's dissertation written at the University of Michigan under the direction of Professor W. B. Ford, and published in pamphlet form by Edwards Brothers of Ann Arbor, Michigan.

2. For purposes of future reference we shall first set down some of the more important relations which occur in obtaining the fundamental solutions

---

* Presented to the American Mathematical Society, April 10, 1936. Received by the Editors, March 4, 1937.

[1] A. R. Forsyth, *A Treatise on Differential Equations*, 3rd ed., London, Macmillan and Co. (1903), pp. 203-222.

[2] Ernst Lindelöf, "Sur l'Intégration de l'Équation de Kummer," *Acta Societatis Fennicae*, (1), Tome 19 (1893), pp. 3-31.

[3] E. W. Barnes, "A new development of the theory of the hypergeometric functions," *Proceedings of the London Mathematical Society*, Series 2, vol. 6 (1908), pp. 141-177.

of (1) about each of its three regular singular points $z=0$, $z=1$, $z=\infty$ by the usual methods of the Fuchs[4] theory.

The indicial equation of (1) corresponding to the singular point $z=0$ is

$$k(k-1)+\gamma k=0, \tag{2}$$

the roots of which are found to be

$$k_1=0, \qquad k_2=1-\gamma. \tag{3}$$

Assuming that k_1-k_2 is non-integral, we shall therefore have, according to the Fuchs theory two fundamental solutions Y_1 and Y_2 about the point $z=0$ having the forms

$$Y_1=z^{k_1}\sum_{n=0}^{\infty}g_1(n)z^n; \qquad Y_2=z^{k_2}\sum_{n=0}^{\infty}g_2(n)z^n. \tag{4}$$

Moreover, if we put

$$f(z,k)=k(k-1)(1-z)+k[\gamma-(\alpha+\beta+1)z]-\alpha\beta z,$$

and

$$f_0(k)=[f(z,k)]_{z=0}=k(k-1)+\gamma k,$$

$$f_1(k)=\left[\frac{\partial f(z,k)}{\partial z}\right]_{z=0}=-k(k-1)-(\alpha+\beta+1)\,k-\alpha\beta,$$

then the same theory indicates that $g_1(n)$ will satisfy the linear recurrence relation

$$g_1(n)f_0(k+n)+g_1(n-1)f_1(k+n-1)=0 \tag{5}$$

in which k is given the value $k_1=0$; while $g_2(n)$ will satisfy the same relation with the value $k_2=1-\gamma$ used for k.

The values of $g_1(0)$, $g_2(0)$ are arbitrary. For definiteness, let us take $g_1(0)=g_2(0)=1$, in which case $g_1(n)$, $g_2(n)$ become completely determined by (5) for $n=1,2,3,\cdots$.

When the $g_1(n)$, $g_2(n)$ are thus determined we find that when $k_1-k_2=\gamma-1$ is non-integral the two fundamental solutions about the singular point $z=0$ can be put in the forms

$$Y_1=F(\alpha,\beta,\gamma;z), \quad Y_2=z^{1-\gamma}F(\alpha-\gamma+1,\beta-\gamma+1,2-\gamma;z), \tag{6}$$

where the $F(\alpha,\beta,\gamma;z)$ represents the well-known Hypergeometric Function defined by

$$F(\alpha,\beta,\gamma;z)=1+\sum_{n=1}^{\infty}\frac{\alpha(\alpha+1)\cdots(\alpha+n-1)\beta(\beta+1)\cdots(\beta+n-1)}{1\cdot2\cdot3\cdots n\gamma(\gamma+1)\cdots(\gamma+n-1)}z^n.$$

---

[4] See J. Horn, *Gewöhnliche Differentialgleichungen beliebiger Ordnung*, Sammlung Schubert L., Leipzig (1905), Section 34.

The indicial equation of (1) corresponding to the point $z = \infty$ is found to be

$$k(k-1) - (\alpha + \beta - 1)\, k + \alpha\beta = 0, \tag{7}$$

the roots of which are

$$k'_1 = \alpha, \qquad k'_2 = \beta, \tag{8}$$

after which we may show that if $k'_1 - k'_2 = \alpha - \beta$ is non-integral the two fundamental solutions of (1) about the singular point $z = \infty$ are

$$\begin{aligned} Y_5 &= z^{-\alpha} F(\alpha, \alpha - \gamma + 1, \alpha - \beta + 1; 1/z), \\ Y_6 &= z^{-\beta} F(\beta, \beta - \gamma + 1, \beta - \alpha + 1; 1/z). \end{aligned} \tag{9}$$

When the exponents (8) differ by an integer, that is when

$$\alpha - \beta = n = \text{an integer} \geq 0, \tag{10}$$

the point $z = \infty$ becomes a so-called logarithmic point. The fundamental solutions about $z = \infty$ then take the forms

$$\begin{aligned} &Y_5 = \text{as defined in } (9), \\ &\bar{Y}_6 = Y_5 \operatorname{Log}(1/z) + z^{-k'_2} \sum_{v=0}^{\infty} \bar{h}_v z^{-v};\ h_n = 0, h_0 \neq 0 \text{ unless } n = 0. \end{aligned} \tag{11}$$

3. It is our purpose in what follows to determine the exact forms of the linear relations which connect Y_1 with Y_5 and $\bar{Y}_6$; likewise those connecting Y_2 with the same Y_5, $\bar{Y}_6$. The methods which we shall use rest upon certain general results obtained by Professor Ford in Chapter I of his recent book entitled *The Asymptotic Developments of Functions Defined by Maclaurin Series.*[5] In particular, we shall employ the following General Theorem there established:

"THEOREM. *If the coefficients $g(n)$ of the power series*

$$f(z) = \sum_{n=0}^{\infty} g(n) z^n;\ \textit{radius of convergence} > 0, \tag{A}$$

may be considered as a function $g(w)$ of the complex variable $w = x + iy$ and as such satisfies the two following conditions when considered throughout any arbitrary right half plane $x > x_0$:

(a) *is single-valued and analytic,*

(b) *is such that for all $|\, y \,|$ sufficiently large one may write*

$$|\, g(x + iy) \,| < K e^{\epsilon |y|} \tag{B}$$

---

[5] *Michigan Science Series,* vol. 11 (1936).

where ϵ is an arbitrarily small positive quantity given in advance and where K depends only upon x_0 and ϵ, then the function $f(z)$ defined by (A) is analytic throughout any sector S (vertex at origin) of the z-plane which does not include the positive half of the real axis and $f(z)$ within S is developable asymptotically as follows:

$$(C) \qquad f(z) \sim -\sum_{n=0}^{\infty} \frac{g(n)}{z^n}."$$

We shall make use also of the following Remarks (b) and (e) relative to the foregoing theorem, as noted in the same work.

"(b) In case conditions (a) and (b) of the theorem are satisfied except that $g(w)$ has p $(p \geqq 1)$ singularities situated at the points $w = w_1, w_2, w_3, \cdots, w_p$, none of which are negative integers, the theorem continues to hold true provided one subtracts from the right member of (C) the sum of p loop integrals of the function

$$(D) \qquad \frac{g(w)(-z)^w}{2i \operatorname{Sin} \pi w}, \cdots.$$

"In case the singular points $w = w_m$ are poles, the loop integrals may evidently be replaced by integrations of (D) over small circles, so that in such cases the theorem continues to hold, provided that one subtracts from the right member of (C) the sum of the residues of the function

$$(E) \qquad \frac{\pi g(w)(-z)^w}{\operatorname{Sin} \pi w}$$

at the various poles $w = w_m$.

"(e) The theorem may be applied to any Maclaurin series (A) in which $g(w)$, besides satisfying condition (a), is such that we may write, when $x > x_0$ and $|w|$ is large, $|g(w)| < K|w|^c$, where K and c are constants of which the latter may be positive, negative or zero."

Whenever we apply Remark (b) it shall be understood that for any given value of $z = \rho e^{i\phi}$ the function $(-z)^w$ is rendered precise in meaning through the following convention:

$$(12) \qquad (-z)^w = e^{w \log(-z)} = e^{w[\log \rho + i(\phi - \pi)]} = z^w e^{-i\pi w}; \qquad 0 \leqq \phi < 2\pi.$$

4. Theorem. Employing the above results, we proceed to establish the following theorem which, so far as we have been able to determine, is new:

THEOREM. *The solutions Y_1 and Y_2 defined in (6), when extended analytically outside their circle of convergence, may be expressed linearly in* *terms of the solutions* Y_5 *and* $\bar{Y}_6$, *defined in* (11), *in the following forms, it being understood throughout that* $0 < \arg z < 2\pi$:

$$Y_1 = \frac{\Gamma(\gamma)[\psi(\alpha) + \psi(\gamma - \alpha) + 2c + i\pi - \sum_1^n (1/s)]e^{i\pi(\beta+1)}}{\Gamma(\beta)\Gamma(\gamma - \alpha)\Gamma(n+1)} Y_5 + \frac{\Gamma(\gamma)e^{i\pi(\beta+1)}}{\Gamma(\beta)\Gamma(\gamma-\alpha)\Gamma(n+1)} \bar{Y}_6$$

$$Y_2 = \frac{\Gamma(2-\gamma)[\psi(\alpha - \gamma + 1) + \psi(1-\alpha) + 2c + i\pi - \sum_1^n (1/s)]e^{i\pi(\beta-\gamma)}}{\Gamma(\beta - \gamma + 1)\Gamma(1-\alpha)\Gamma(n+1)} Y_5 + \frac{\Gamma(2-\gamma)e^{i\pi(\beta-\gamma)}}{\Gamma(\beta-\gamma+1)\Gamma(1-\alpha)\Gamma(n+1)} \bar{Y}_6$$

where $\psi(\alpha)$ *is the well-known Psi-function defined as the logarithmic derivative of the Gamma-function, that is* $\psi(\alpha) = \Gamma'(\alpha)/\Gamma(\alpha)$, *where* C *denotes Euler's constant defined by* $C = -\psi(1)$, *and where* $\sum_1^n (1/s)$ *becomes* 0 *when* $n = 0$.[6]

Proof. Returning to the recurrence relation (5), we substitute the values of $f_0(k+n)$ and $f_1(k+n-1)$ and, in order that the notation shall conform to that which is customary in the theory of difference equations, we replace $g(n)$, $g(n-1)$ respectively by $u(x)$, $u(x-1)$ and subsequently advance x to $x+1$. The relation (5) thus becomes

$$(13) \qquad q_1(x)u(x+1) + q_0(x)u(x) = 0,$$

where

$$q_1(x) = (k+x+1)(k+x) + \gamma(k+x+1),$$
$$q_0(x) = -(k+x)(k+x-1) - (\alpha+\beta+1)(k+x) - \alpha\beta,$$

and k takes the value k_1 or k_2 of (3).

For those particular values of k in which we are interested, namely $k = k_1$ and $k = k_2$ of (3), the roots of $q_1(x) = 0$ are found to be

$$x = -1, \qquad x = -(2k+\gamma),$$

and the roots of $q_0(x) = 0$ are

$$(14) \qquad r_1 = -\alpha - k, \qquad r_2 = -\beta - k.$$

Hence we may write

$$q_1(x) = (x+1)(x+2k+\gamma), \qquad q_0(x) = -(x - r_1)(x - r_2).$$

---

[6] See L. M. Milne-Thomson, *The Calculus of Finite Differences*, London, Macmillan and Co. (1933), p. 245 and p. 250.

When we substitute these values into (13) and then solve the resulting first order difference equation by elementary methods we obtain as the particular solution $u(x)$ which takes the value 1 when $x=0$

$$u(x)=g(x)=\frac{\Gamma(2k+\gamma)\Gamma(x-r_1)\Gamma(x-r_2)}{\Gamma(-r_1)\Gamma(-r_2)\Gamma(x+1)\Gamma(x+2k+\gamma)}, \tag{15}$$

in which we shall assume at first that neither r_1 nor r_2 is zero or a positive integer, thus rendering $u(x)\not\equiv 0$.

We must notice here that since $\alpha-\beta=n=$ an integer $\geqq 0$, it follows from (14) that $r_1=r_2-n$.

Now, each of the fundamental solutions Y_1, Y_2 is of the form

$$Y=z^k\sum_{n=0}^{\infty}g(n)z^n;\qquad g(0)=1, \tag{16}$$

where the $g(n)$ is defined by (15). The function $g(x)$ in (15) coincides with the $g(n)$ of (16) when $x=0,1,2,3,\cdots$ and is, moreover, analytic throughout the finite x-plane except for poles of the first order at the points

$$x=r_2, r_2-1, r_2-2,\cdots, r_2-n+1 \tag{17}$$

and poles of the second order at

$$x=r_1, r_1-1, r_1-2,\cdots, \tag{18}$$

while in distant portions of the plane lying in any arbitrary right half-plane, the same function $g(x)$ satisfies the condition described in Remark (e) quoted above.

In order to determine the asymptotic behavior of the function $Y_1(x)$ defined by the Maclaurin series (16) we may therefore apply the General Theorem quoted above, subject to Remark (b). When we do so, we see that in our present case the $g(-x)$; $x=1,2,\cdots$, in formula (C) all vanish owing to the general relation $1/\Gamma(-x)=0$; $x=1,2,3,\cdots$. Furthermore, Remark (b) requires that we subtract from the right-hand side of (C) the sum of the residues of the function

$$\begin{aligned}\frac{\pi g(x)(-z)^x}{\operatorname{Sin}\pi x}&=\Gamma(x)\Gamma(1-x)g(x)(-z)^x\\&=-\frac{\Gamma(-x)\Gamma(2k+\gamma)\Gamma(x-r_1)\Gamma(x-r_2)}{\Gamma(-r_1)\Gamma(-r_2)\Gamma(x+2k+\gamma)}(-z)^x\end{aligned} \tag{19}$$

at its poles (17) and (18). However, inasmuch as we are interested only in determining the values of the constants which join the solutions Y_1 and Y_2

linearly with the solutions Y_5 and Y_6 it is sufficient for our purpose to determine the residues of (19) at the two poles $x = r_1$ and $x = r_2$.

The residue of (19) at the first of the points (17) involves z to the power $-\beta - k$ and the residues at the remaining points of the same set involve z to the lower powers $-\beta - k - 1$ to $-\beta - k - n + 1 = -\alpha - k + 1$. It will be shown presently that the residue at the first of the poles (18) involves $z^{-\alpha-k}$ and also $z^{-\alpha-k} \log z$, while the residues at the remaining points (18) involve $z^{-\alpha-k-s}$ and $z^{-\alpha-k-s} \log z$ $(s = 1, 2, 3, \cdots)$. It follows that the coefficient of the highest power of z from both sources will come from the residue at the first point of (17). In order to determine the linear relations which we are seeking it suffices then to determine the residue of (19) at the point $x = r_2$ and the logarithmic part of the residue of (19) at the point $x = r_1$. The residue at the pole $x = r_2$ is found by an elementary theorem in the calculus of residues to be

$$-\frac{\Gamma(2k + \gamma)\Gamma(r_2 - r_1)}{\Gamma(-r_1)\Gamma(r_2 + 2k + \gamma)} (-z)^{r_2}. \tag{20}$$

When we substitute into (20) the values of r_1 and r_2 as defined, and apply the convention (12) to $(-z)^{r_2}$ this residue may be written as

$$- C_2(k, \alpha, \beta, \gamma) z^{-\beta-k} = - \frac{\Gamma(2k + \gamma)\Gamma(n) e^{i\pi(\beta+k)} z^{-\beta-k}}{\Gamma(\alpha + k)\Gamma(\gamma - \beta + k)}. \tag{21}$$

In order to obtain the residue of (19) at its pole of the second order $x = r_1$ we write (19) in the form $-\dfrac{M(x) z^x}{Q^2(x)}$, where

$$M(x) = \frac{\Gamma(-x)\Gamma(2k+\gamma)\Gamma^2(x-r_1+1) e^{-i\pi x}}{\Gamma(-r_1)\Gamma(-r_2)\Gamma(x+2k+\gamma)(x-r_1-1)(x-r_1-2)\cdots(x-r_1-n)},$$

$$Q(x) = x - r_1, \tag{22}$$

The residue of (19) at $x = r_1$ is now found to be

$$- M'(r_1) z^{r_1} - M(r_1) z^{r_1} \log z. \tag{23}$$

As we have explained above, we have at first only to determine the logarithmic part of this residue, which by reference to (22) is easily found to be

$$- C_1(k, \alpha, \beta, \gamma) z^{-\alpha-k} \log z = \frac{\Gamma(2k + \gamma) e^{i\pi(\beta+k)} z^{-\alpha-k}}{\Gamma(\beta + k)\Gamma(k + \gamma - \alpha)\Gamma(n + 1)} \log z. \tag{24}$$

We may now apply the General Theorem previously quoted together with its accompanying remarks. In this way we arrive at the following preliminary result:

The function Y defined in (16), when extended analytically outside its circle of convergence, may be developed asymptotically in the following form, it being understood throughout that $0 < \arg z < 2\pi$:

$$(25) \qquad Y \sim C_1(k, \alpha, \beta, \gamma) z^{-\alpha}[1 + (\ )/z + (\ )/z^2 + \cdots] \log z \\ + C_2(k, \alpha, \beta, \gamma) z^{-\beta}[1 + (\ )/z + (\ )/z^2 + \cdots]$$

where $C_1(k, \alpha, \beta, \gamma)$ and $C_2(k, \alpha, \beta, \gamma)$ are defined in (24) and (21) respectively. The asymptotic expansion of the solution Y_1 of (1) results from this when the value $k = k_1 = 0$ is used, while the corresponding expansion for Y_2 results likewise from the use of $k_2 = 1 - \gamma$.

We are concerned here, however, with the expression of the function Y_1 as a linear combination of the solutions Y_5 and $\bar{Y}_6$, that is, we wish to determine constants K_1 and K_2 such that

$$Y_1 \sim K_1 Y_5 + K_2 \bar{Y}_6$$

with a similar expression for Y_2. Upon referring to the definitions (11) of Y_5 and $\bar{Y}_6$, we can identify by (25) the constant K_2 with $-C_1(k, \alpha, \beta, \gamma)$. The solution Y_5 contains the factor $z^{-\alpha} = z^{-\beta-n}$. We know from the Fuchs theory that the non-logarithmic part of the solution $\bar{Y}_6$ does not involve z to this power since $h_n = 0$. We may therefore identify the constant K_1 with the coefficient of $z^{-\beta-n}$ in the non-logarithmic part of the right side of (25). This coefficient, before taking out the factor $C_2(k, \alpha, \beta, \gamma)$, is the value of the coefficient of $z^{r_1} = z^{-\alpha-k}$ in the non-logarithmic part of the residue (23). This is easily found, by use of (22), to be

$$(26) \quad C_n(k, \alpha, \beta, \gamma) \\ = \frac{\Gamma(2k+\gamma)[\psi(\alpha+k) + \psi(\gamma-\alpha+k) + 2C + i\pi - \sum_1^n (1/s)] e^{i\pi(\beta+k+1)}}{\Gamma(\beta+k)\Gamma(\gamma-\alpha+k)\Gamma(n+1)}.$$

We thus arrive at the following result for the solution Y_1:

The solution Y_1, when extended analytically outside its circle of convergence, may be developed asymptotically in the following form, it being understood throughout that $0 < \arg z < 2\pi$:

$$(27) \qquad Y_1 \sim C_n(k, \alpha, \beta, \gamma) Y_5 - C_1(k, \alpha, \beta, \gamma) \bar{Y}_6$$

where $C_n(k, \alpha, \beta, \gamma)$ and $C_1(k, \alpha, \beta, \gamma)$ are defined in (26) and (24) respectively and in which k takes the value $k_1 = 0$.

The corresponding result expressing the solution Y_2 linearly in terms of Y_5 and $\bar{Y}_6$ is obtained from the foregoing by the use of $k = k_2 = 1 - \gamma$ for k.

We note here that the constants $C_1(k, \alpha, \beta, \gamma)$, $C_2(k, \alpha, \beta, \gamma)$ and

$C_n(k, \alpha, \beta, \gamma)$ all preserve meanings regardless of whether $r_1 = -\alpha - k$ or $r_2 = -\beta - k$ is zero or a positive integer, so that the restriction made in (15) may now be removed. Since the series Y_5 and $\bar{Y}_6$ are known to be convergent for the indicated values of z, the symbol $\sim$ may be changed to $=$. We have now only to introduce into (27) the values of the constants $C_n(k, \alpha, \beta, \gamma)$ and $C_1(k, \alpha, \beta, \gamma)$ as defined in order to arrive at the final results stated in the Theorem.

5. Further results. The solutions of (1) about the point $z = 1$ are

$$Y_3 = \mathrm{F}(\alpha, \beta, \alpha + \beta - \gamma + 1; 1 - z),$$
$$Y_4 = (1 - z)^{\gamma-\alpha-\beta} F(\gamma - \alpha, \gamma - \beta, \gamma - \alpha - \beta + 1; 1 - z).$$

In order to determine the linear relationships between the solutions Y_3 and Y_4 and the solutions Y_5 and $\bar{Y}_6$ we first make the transformation $z = 1 - z'$ in (1) and subsequently apply the same methods to this transformed equation as we have employed in the determination of the results in Theorem I. After these results have been obtained, we make the reverse transformation $z' = 1 - z$. The final result obtained in this manner is stated in the following theorem:

THEOREM II. *The solutions Y_3 and Y_4, when extended analytically outside their circle of convergence, may be expressed linearly in terms of Y_5 and $\bar{Y}_6$ in the following forms, it being understood throughout that* $0 < \arg\,(1 - z) < 2\pi$:

$$Y_3 = \frac{\Gamma(\alpha+\beta-\gamma+1)\left[\psi(\alpha)+\psi(\beta-\gamma+1)+2c+i\pi-\sum_1^n(1/s)\right]e^{i\pi(\beta+1)}}{\Gamma(\beta)\Gamma(\beta-\gamma+1)\Gamma(n+1)}\,Y_5$$
$$+\frac{\Gamma(\alpha+\beta-\gamma+1)e^{i\pi(\beta+1)}}{\Gamma(\beta)\Gamma(\beta-\gamma+1)\Gamma(n+1)}\,\bar{Y}_6$$

$$Y_4 = \frac{\Gamma(\gamma-\alpha-\beta+1)\left[\psi(\gamma-\beta)+\psi(1-\alpha)+2c+i\pi-\sum_1^n(1/s)\right]e^{i\pi(\gamma-\alpha+1)}}{\Gamma(\gamma-\alpha)\Gamma(1-\alpha)\Gamma(n+1)}\,Y_5$$
$$+\frac{\Gamma(\gamma-\alpha-\beta+1)e^{i\pi(\gamma-\alpha+1)}}{\Gamma(\gamma-\alpha)\Gamma(1-\alpha)\Gamma(n+1)}\,\bar{Y}_6$$

When the fundamental solutions about either the singular point $z = 0$ or the singular point $z = 1$ are logarithmic the resulting interrelations are obtained by first making the appropriate transformation of the independent variable in (1), employing the same procedure as we have used above in the proof of Theorem I, and then making the reverse transformation in the results.

ARIZONA STATE TEACHERS
COLLEGE AT FLAGSTAFF.

THE THEOREMS OF GAUSS-BONNET AND STOKES.*

By E. R. VAN KAMPEN.

The idea of a parallel displacement in vector analysis opened new possibilities also for the differential geometry of surfaces. The present note contains an elementary proof of the Gauss-Bonnet theorem based on this idea. A systematic use is made of the correspondence between the surface and the spaces of its parameter systems for purposes of subdivision in **7** and for the determination of the variation of an angle in **4**.

Third continuous derivatives of the parametric equations of the surface occur only at the beginning of **5**. Their use may be eliminated by the short additional consideration in **6**. Thus one obtains also for the *Theorema Egregium* (without the explicit expression for the Gaussian curvature) a proof for surfaces of class C_2. It may be of interest that the proof below makes no use either of the second or of the first fundamental form.

In **10** a short proof of Stokes theorem is given which is valid for a vector-field of class C_1 and a compact region T on a surface of class C_1 if the boundary of T consists of rectifiable arcs.

1. Let the letters $\boldsymbol{x}, \boldsymbol{u}, \boldsymbol{v}, \boldsymbol{w}, \cdots$ represent vectors in 3-space and let $\boldsymbol{u} \cdot \boldsymbol{v}$, $\boldsymbol{u} \times \boldsymbol{v}$, $(\boldsymbol{uvw}) = \boldsymbol{u} \cdot (\boldsymbol{v} \times \boldsymbol{w})$ represent the scalar product, vector product and triple product of vectors respectively. Let a surface S be given by a parametric representation of the form

$$\boldsymbol{x} = \boldsymbol{x}(u^1, u^2), \tag{1}$$

defined in the whole plane of the scalar parameters u^1, u^2. The letters R and T will be used to designate curves and regions of S and at the same time the corresponding curves and regions of the u^i-plane. The label i will have the range 1, 2 and any term containing such a label both as a subscript and as a superscript must be summed over this range. It will be supposed that (1) is at least of class C_1 and that

$$(\boldsymbol{x}_1 \times \boldsymbol{x}_2) \neq 0, \tag{2}$$

where the subscripts 1 and 2 represent partial differentiation with respect to

---

* Received November 15, 1937.

u^1 and u^2 respectively. The normal vector of S is defined as the vector (2) divided by its length and will be denoted by $\boldsymbol{x}_3$. Thus one has

$$(3) \qquad (\boldsymbol{x}_1\boldsymbol{x}_2\boldsymbol{x}_3) > 0 \quad \text{and} \quad \boldsymbol{x}_3(\boldsymbol{x}_1\boldsymbol{x}_2\boldsymbol{x}_3) = (\boldsymbol{x}_1 \times \boldsymbol{x}_2).$$

It is well known that the area of a region T of S may be represented in the form

$$(4) \qquad \int_T\!\!\int (\boldsymbol{x}_1\boldsymbol{x}_2\boldsymbol{x}_3)\, du^1du^2 = \int_T\!\!\int d\sigma.$$

Similarly, if

$$(5) \qquad \boldsymbol{x} = \boldsymbol{x}_3(u^1, u^2)$$

represents, in case of a surface S of class C_2, the spherical image of S, one obtains for the area of the spherical image of a region T of S

$$(6) \qquad \int_T (\boldsymbol{x}_{31}\boldsymbol{x}_{32}\boldsymbol{x}_3)\, du^1du^2,$$

where the usual conventions have to be used if the correspondence between T and its image on the sphere is not one-to-one. One may define the Gaussian curvature K of S at a point P of S as the limit of the quotient of the area of T and of its spherical image, when T is a variable part of S of simple form (e. g. a u^i-square) which has P as limit. One finds, from (4) and (6)

$$(7) \qquad (\boldsymbol{x}_{31}\boldsymbol{x}_{32}\boldsymbol{x}_3) = K(\boldsymbol{x}_1\boldsymbol{x}_2\boldsymbol{x}_3).$$

This well-known formula may be taken immediately from the usual definition of K by means of the Weingarten differentiation formulae.

2. Let an arc R of class C_2 on a surface S of class C_2 be represented in the form

$$(8) \qquad u^i = u^i(s),$$

where s is the arc length of R. On substituting (8) in (1) and differentiating, one obtains

$$(9) \qquad \dot{\boldsymbol{x}} = \boldsymbol{x}_i\, \dot{u}^i,$$

where the $\dot{}$ denotes differentiation with respect to s and the tangent vector $\dot{\boldsymbol{x}}$ of R has length 1 since the arc length is parameter on R.

It is clear that there is a one-to-one correspondence between vectors in the u^i-plane which are attached at a certain point and vectors tangent to S at the corresponding point of S and that this correspondence is continuous both ways. For instance to (9) there corresponds the vector $\dot{u}^i$ attached at the point (8).

A system

$$v = v(s) \tag{10}$$

of vectors which is of class C_1 and is tangent to S at the point (8) of S is called *parallel along R* if

$$\dot{v} = \dot{v}(s) \text{ is normal to } S \text{ at } (8). \tag{11}$$

On placing $\boldsymbol{v} = v^i \boldsymbol{x}_i$ and $\dot{\boldsymbol{v}} \cdot \boldsymbol{x}_i = 0$ one obtains linear differential equations for the $v^i(s)$, so that a system (10) exists and is uniquely determined if the $v^i(s)$ are given for one value of s. Since $\dot{\boldsymbol{v}}$ is perpendicular to $\boldsymbol{v}$, the length of $\boldsymbol{v}$ is independent of s. *If* (10) *is parallel along R, then so is any vector obtained from* (10) *by rotation of* (10) *in the tangent plane of S at* (8) *over an angle which is independent of s.* For if (10) is parallel, so is $\boldsymbol{w} = \boldsymbol{x}_3 \times \boldsymbol{v}$ which is obtained from $\boldsymbol{v}$ by a rotation over the angle $\frac{1}{2}\pi$; in fact, $\dot{\boldsymbol{w}} = \dot{\boldsymbol{x}}_3 \times \boldsymbol{v} + \boldsymbol{x}_3 \times \dot{\boldsymbol{v}}$ is normal to S at (8), since $\dot{\boldsymbol{x}}_3$ is tangent and $\dot{\boldsymbol{v}}$ is normal. Furthermore if any two vectors are parallel along R, so are their linear combinations with constant coefficients.

It follows that, if ϑ is the angle from (10) to (9), then $\dot{\vartheta}$ is independent of the particular system (10) chosen. The derivative $\dot{\vartheta}$ is called the *geodesic* curvature of R and will be denoted by κ. One finds, if $\boldsymbol{v}$ has length 1,

$$\sin\vartheta = (\boldsymbol{v}\dot{\boldsymbol{x}}\boldsymbol{x}_3), \qquad \cos\vartheta = \boldsymbol{v} \cdot \dot{\boldsymbol{x}}, \tag{12}$$

hence

$$\dot{\vartheta}(\boldsymbol{v} \cdot \dot{\boldsymbol{x}}) = (\dot{\boldsymbol{v}}\dot{\boldsymbol{x}}\boldsymbol{x}_3) + (\boldsymbol{v}\ddot{\boldsymbol{x}}\boldsymbol{x}_3) + (\boldsymbol{v}\dot{\boldsymbol{x}}\dot{\boldsymbol{x}}_3).$$

Here the first and third triple products are 0, since $\dot{\boldsymbol{v}}$ is normal and $\dot{\boldsymbol{x}}_3$ is tangent to S. On choosing $\boldsymbol{v}$ to be $\dot{\boldsymbol{x}}$ at the point of R under consideration, one finds $\boldsymbol{v} \cdot \dot{\boldsymbol{x}} = 1$, hence

$$\kappa = \dot{\vartheta} = (\dot{\boldsymbol{x}}\ddot{\boldsymbol{x}}\boldsymbol{x}_3). \tag{13}$$

3. Let a vector $\boldsymbol{w} = \boldsymbol{w}(s)$ be tangent to S at the point (8), let $\boldsymbol{w}$ be of class C_1 and of length 1 and let the angles from $\boldsymbol{w}$ to $\boldsymbol{x}$ and to $\boldsymbol{v}$ be denoted by ω and ϕ. Then one has

$$\omega = \vartheta + \phi, \tag{14}$$

if care is taken that all three angles are continuous and that (14) holds exactly (and not only modulo 2π) at some point of R. Furthermore

$$\sin\phi = (\boldsymbol{w}\boldsymbol{v}\boldsymbol{x}_3), \qquad \cos\phi = \boldsymbol{w} \cdot \boldsymbol{v}, \tag{15}$$

and so, in the same way as (13) from (12),

$$\dot{\phi} = (\dot{\boldsymbol{w}}\boldsymbol{w}\boldsymbol{x}_3). \tag{16}$$

Now let an arc R be divided by a finite number of vertices P_k into arcs of class C_2. Denote by α_k the angle from $\dot{\boldsymbol{x}}(s_k - 0)$ to $\dot{\boldsymbol{x}}(s_k + 0)$, where $-\pi \leqq \alpha_k \leqq \pi$ and s_k is the value of s at P_k. If $\alpha_k = \pm\pi$ choose the signature of α_k in such a way that at every P_k the angle subtended by the region to the right of R is $\alpha_k + \pi$.

Let $\boldsymbol{w}$ be continuous at P_k and of class C_1 on the remainder of R. Clearly a parallel system $\boldsymbol{v}$ may be defined along R having the same properties. If the convention is used that

$$\vartheta(s_k + 0) - \vartheta(s_k - 0) = \alpha_k, \qquad \omega(s_k + 0) - \omega(s_k - 0) = \alpha_k, \tag{17}$$

while ϕ is continuous for every s, then (14) remains true for the arc R. If $\Delta\omega, \Delta\vartheta, \Delta\phi$ denotes the increase of ω, ϑ, ϕ, while s increases from its initial value to its final value, one has, by (14) and (17),

$$\Delta\omega = \Delta\vartheta + \Delta\phi \tag{18}$$

and one obtains from (13) and (17) and from (16)

$$\Delta\vartheta = \int_R \kappa\, ds + \Sigma_k \alpha_k \quad \text{and} \quad \Delta\phi = \int_R (\dot{\boldsymbol{w}}\boldsymbol{w}\boldsymbol{x}_3)\, ds, \tag{19}$$

hence, from (18),

$$\Delta\omega = \int_R \kappa\, ds + \int_R (\dot{\boldsymbol{w}}\boldsymbol{w}\boldsymbol{x}_3)\, ds + \Sigma_k \alpha_k. \tag{20}$$

4. The meaning of (20) is not lost if R is a simple closed curve and the parameter s begins and ends at a point of R where the second derivative exists. It will be assumed for the present that R is a simple closed curve which is the positively oriented boundary of a compact portion T of S. In that case it is easy to define a vector field $\boldsymbol{w}$ which is of class C_1 on T, hence continuous on R and of class C_1 on each arc of class C_2 of R. One can for instance transfer a suitable field from the u^i-plane to S. It will be shown that in the case considered

$$\Delta\omega = 2\pi \tag{21}$$

First, $\Delta\omega = \Delta\lambda$, where λ is the angle in the u^i-plane, from the vector w^i which corresponds to $\boldsymbol{w}$ to the vector $\dot{u}^i$ which corresponds to $\dot{\boldsymbol{x}}$. This is clear, since both $\Delta\omega$ and $\Delta\lambda$ are integer multiples of 2π, while, for increasing s, the angles ω and λ both pass at the same time and in the same direction through a multiple of π.

Next,[1] $\Delta\lambda = \Delta\mu$, where μ is the inclination of u^i, i. e. the angle from the vector $(1, 0)$ to the vector (u^1, u^2). Note that T, which is compact by assumption, is represented in the u^i-plane by the simple closed curve R together with its interior. Now on this set the inclination of w^i may be determined as a continuous function.[2] Hence w^i may be changed into the constant vector field $(1, 0)$ by a continuous rotation. This proves $\Delta\lambda = \Delta\mu$, since both $\Delta\lambda$ and $\Delta\mu$ are integral multiples of 2π.

In order to prove (21) it remains to show that $\Delta\mu = 2\pi$. Now this is the "Umlaufsatz" of which simple proofs may be found in the literature.[3]

As a consequence of (21), (20) takes the form

$$\int_R \kappa \, ds + \int_R (\dot{\boldsymbol{w}}\boldsymbol{w}\boldsymbol{x}_3) \, ds = 2\pi - \Sigma_k \alpha_k. \tag{22}$$

5. Let it be assumed for a moment that S is of class C_3, so that $\boldsymbol{w}$ may be chosen of class C_2. Then the theorem of Green may be applied to the second integral in (22) as follows,

$$\begin{aligned} \int_R (\dot{\boldsymbol{w}}\boldsymbol{w}\boldsymbol{x}_3) \, ds &= \int_R [(\boldsymbol{w}_1\boldsymbol{w}\boldsymbol{x}_3) \, du^1 + (\boldsymbol{w}_2\boldsymbol{w}\boldsymbol{x}_3) \, du^2] \\ &= \int\!\!\int_T [(\boldsymbol{w}_2\boldsymbol{v}\boldsymbol{x}_3)_1 - (\boldsymbol{w}_1\boldsymbol{w}\boldsymbol{x}_3)_2] \, du^1 du^2. \end{aligned} \tag{23}$$

Now one has for the integrand of the last integral,

$$\begin{aligned} (\boldsymbol{w}_2\boldsymbol{w}\boldsymbol{x}_3)_1 - (\boldsymbol{w}_1\boldsymbol{w}\boldsymbol{x}_3)_2 &= (\boldsymbol{w}_{21}\boldsymbol{w}\boldsymbol{x}_3) \\ &+ (\boldsymbol{w}_2\boldsymbol{w}_1\boldsymbol{x}_3) + (\boldsymbol{w}_2\boldsymbol{w}\boldsymbol{x}_{31}) - (\boldsymbol{w}_{12}\boldsymbol{w}\boldsymbol{x}_3) - (\boldsymbol{w}_1\boldsymbol{w}_2\boldsymbol{x}_3) - (\boldsymbol{w}_1\boldsymbol{w}\boldsymbol{x}_{32}). \end{aligned} \tag{24}$$

---

[1] It is clear that this step in the argument could be avoided by choosing w in a suitable way to begin with.

[2] It is possible to subdivide T into arbitrarily small polygons with edges of class C_2. One may for instance use line segments in the u^i-plane. If μ could not be defined as a continuous function on T, the variation of μ along the boundary of at least one of these polygons would be $2\pi n \neq 0$. But this is clearly impossible if the polygons are sufficiently small.

[3] H. Hopf, *Compositio Mathematica*, vol. 2 (1935), p. 50 and pp. 53-55, where further references are given; also E. R. van Kampen, *Compositio Mathematica*, vol. 4 (1937), p. 272.

Here the first and fourth terms on the right cancel, while the second and fifth are both zero, since they contain only vectors perpendicular to $\boldsymbol{w}$. Hence

$$(25)\qquad \int_R (\dot{\boldsymbol{w}}\boldsymbol{w}\boldsymbol{x}_3)\,ds = \int_T\!\!\int [(\boldsymbol{w}_2\boldsymbol{w}\boldsymbol{x}_{31}) - (\boldsymbol{w}_1\boldsymbol{w}\boldsymbol{x}_{32})]\,du^1du^2.$$

It remains to evaluate the integrand of the last integral. Now $\boldsymbol{w}$ and $\boldsymbol{x}_{31}$ are tangent to S, so that in $(\boldsymbol{w}_2\boldsymbol{w}\boldsymbol{x}_{31})$, $\boldsymbol{w}_2$ may be replaced by its normal component $(\boldsymbol{w}_2\cdot\boldsymbol{x}_3)\boldsymbol{x}_3$ which is equal to $-(\boldsymbol{w}\cdot\boldsymbol{x}_{32})\boldsymbol{x}_3$, since $\boldsymbol{w}\cdot\boldsymbol{x}_3 = 0$. Thus one finds for the integrand on the right of (25),

$$(26)\qquad (\boldsymbol{x}_3\boldsymbol{w}\boldsymbol{x}_{32})(\boldsymbol{w}\cdot\boldsymbol{x}_{31}) - (\boldsymbol{x}_3\boldsymbol{w}\boldsymbol{x}_{31})(\boldsymbol{w}\cdot\boldsymbol{x}_{32}).$$

Applying the identity $\boldsymbol{a}\times(\boldsymbol{b}\times\boldsymbol{c}) = (\boldsymbol{a}\cdot\boldsymbol{c})\boldsymbol{b} - (\boldsymbol{a}\cdot\boldsymbol{b})\boldsymbol{c}$ in two different ways to $\boldsymbol{a}\cdot((\boldsymbol{b}\times\boldsymbol{c})\times(\boldsymbol{d}\times\boldsymbol{e}))$ one obtains

$$\boldsymbol{a}\cdot\boldsymbol{b}(\boldsymbol{cde}) - \boldsymbol{a}\cdot\boldsymbol{c}(\boldsymbol{deb}) + \boldsymbol{a}\cdot\boldsymbol{d}(\boldsymbol{ebc}) - \boldsymbol{a}\cdot\boldsymbol{e}(\boldsymbol{bcd}) = 0.$$

Substituting here $\boldsymbol{w}, \boldsymbol{w}, \boldsymbol{x}_3, \boldsymbol{x}_{31}, \boldsymbol{x}_{32}$ instead of $\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c}, \boldsymbol{d}, \boldsymbol{e}$, one finds that (26) equals $(\boldsymbol{x}_{31}\boldsymbol{x}_{32}\boldsymbol{x}_3)$. Thus (26) is $K(\boldsymbol{x}_1\boldsymbol{x}_2\boldsymbol{x}_3)$ by (7), and (4) shows that (25) takes the form

$$(27)\qquad \int_R (\dot{\boldsymbol{w}}\boldsymbol{w}\boldsymbol{x}_3)\,ds = \int_T\!\!\int K\,d\sigma,$$

so that the theorem of Gauss-Bonnet (in the simplest case) is obtained from (22) in the form:

$$(28)\qquad \int_R \kappa\,ds + \int_T\!\!\int K\,\mathrm{d}\sigma = 2\pi - \Sigma_k\alpha_k.$$

The formula

$$(\boldsymbol{x}_1\boldsymbol{x}_2\boldsymbol{x}_3)K = (\boldsymbol{w}_2\boldsymbol{w}\boldsymbol{x}_3)_1 - (\boldsymbol{w}_1\boldsymbol{w}\boldsymbol{x}_3)_2,$$

which is implicit in (23) and (27) goes over into well-known formulae, if one chooses $\boldsymbol{w}$ to be parallel either to $\boldsymbol{x}_1$ or to $\boldsymbol{x}_2$. It is obvious that the identity of (26) and $(\boldsymbol{x}_{31}\boldsymbol{x}_{32}\boldsymbol{x}_3)$ may be obtained from the rule for $\sin(\alpha-\beta)$ where α and β are the angles from $\boldsymbol{x}_{31}$ and $\boldsymbol{x}_{32}$ to $\boldsymbol{w}$.

6. Now it will be shown that (25) holds for a vector field $\boldsymbol{w}$ of class C_1. If $a = a(u^1, u^2)$ and $b = b(u^1, u^2)$ are scalar functions on T of class C_1, then

$$(29)\qquad \int a\dot{b}\,ds = \int_R (ab_1\,du^1 + ab_2\,du^2) = \int_T\!\!\int (a_1b_2 - a_2b_1)\,du^1du^2.$$

This is a consequence of the theorem of Green, since b, b_1 and b_2 may be approximated uniformly on T by a polynomial in u^1 and u^2 and its two partial derivatives. On replacing b in (29) by the different components of $\boldsymbol{w}$, one may verify that

$$\int_R (\boldsymbol{w} \times \boldsymbol{x}_3) \cdot \dot{\boldsymbol{w}} ds = \int\int_T [(\boldsymbol{w} \times \boldsymbol{x}_3)_1 \cdot \boldsymbol{w}_2 - (\boldsymbol{w} \times \boldsymbol{x}_3)_2 \cdot \boldsymbol{w}_1] du^1 du^2,$$

which identity is equivalent to (25), since $(\boldsymbol{w}_1 \boldsymbol{w}_2 \boldsymbol{x}_3) = 0$.

Since a vector field $\boldsymbol{w}$ of class C_1 always exists on T, provided that S is of class C_2, it follows that (28) holds if S is of class C_2, while R consists of arcs of class C_2.

Since, as is well known, on a surface of class C_2 the geodesic curvature of an arc R depends on the first fundamental form of S only, and since (28) holds for every T, it follows that the Gaussian curvature of a surface S of class C_2 does not depend on the second fundamental form of S. Hence the *Theorema Egregium* holds for surfaces of class C_2.

7. It is, of course, not possible to represent every surface S by one representation (1). However, any surface S may be defined by means of a finite or enumerable collection of representations (1), with the understanding that whenever a part of S is defined by two representations (1), there exists an orientation preserving continuous one-to-one correspondence between the two sets of parameters involved. The parameter representations (1) and the above mentioned correspondence are all supposed to be of the same class C_n, which is also the class of the surface S. Now, *any compact subset T of S, which has as boundary R a graph consisting of arcs of class C_n, may be subdivided by additional arcs of class C_n into a finite number of regions T_m, $m = 1, 2, \cdots$, such that the boundary R_m of T_m is a simple closed curve and that T_m is defined by means of only one of the representations* (1) *of S*. The proof may be sketched as follows. First, as a consequence of the compactness of T, this set is contained in the part of S defined by means of a finite number of representations (1). Next, the part of T defined by a fixed representation (1) is a subset of the corresponding u^i-plane of which the boundary consists of certain arcs of class C_n. The part of T which is not defined by any other representation (1) and not yet subdivided in the desired way is bounded in this u^i-plane. Clearly a somewhat larger part of T may be subdivided in the desired way, for instance by means of line segments. On repeating this for the finite number of representations (1) which are used to define T, one clearly obtains a proof of the above statement.

If the numbers of vertices, edges, regions T_m, which occur in the above subdivision is a_0, a_1, a_2, one terms $C = a_0 - a_1 + a_2$ the *characteristic of* T. In case $n = 2$, the independence of C from the particular subdivision is an automatic consequence of the Gauss-Bonnet theorem.

8. Now suppose that a region T on a surface S of class C_2 has a boundary R consisting of arcs of class C_2 and that a_0, a_1, a_2, T_m, R_m, $m = 1, \cdots, a_2$, have the meaning of **7.** Denote by β_{mk} the *interior* angles of T_m, so that, since (28) holds for each T_m,

$$\int_{R_m} \kappa \, ds + \int\int_{T_m} K \, d\sigma = 2\pi - \Sigma_k(\pi - \beta_{mk}). \tag{30}$$

On taking the sum of (30) for all m one obtains

$$\int_{R} \kappa \, ds + \int\int_{T} K \, d\sigma = 2\pi a_2 - \Sigma_m\Sigma_k(\pi - \beta_{mk}). \tag{31}$$

Let β_k, $k = 1, 2, \cdots$, denote all interior angles of the complement of T on S, so that $\alpha_k = \beta_k - \pi$ are the discontinuities in the tangent vector of R if one follows all boundaries of the complement of T in their negative direction. The α_k will be called the oriented angles of R. The right side of (31) is

$$2\pi a_2 - \Sigma_k\alpha_k - \Sigma_m\Sigma_k(\pi - \beta_{mk}) - \Sigma_k(\pi - \beta_k),$$

or

$$2\pi a_2 - \Sigma_k\alpha_k + (\Sigma_m\Sigma_k\beta_{mk} + \Sigma_k\beta_k) - (\Sigma_m\Sigma_k\pi + \Sigma_k\pi). \tag{32}$$

Here the third term represents the sum of all angles subtended at all vertices of the subdivisions of T used, hence it is equal to $2\pi a_0$. The fourth term contains the number π each time a region T_m (or the complement of T) adjoins a vertex. Hence it also contains the number π each time an edge adjoins a vertex. Thus it is equal to $2\pi a_1$. On using the definition of the characteristic C of T, one sees that (31) goes over into

$$\int_{R} \kappa \, ds + \int\int_{T} K \, d\sigma = 2\pi C - \Sigma_k\alpha_k. \tag{33}$$

Thus one may formulate the general

THEOREM OF GAUSS-BONNET. *If K and $d\sigma$ are Gaussian curvature and area element of a surface S of class C_2, T is a compact region of S with char-*

acteristic C and of which the boundary R consists of a finite number of arcs of class C_2; if furthermore κ is the geodesic curvature and ds the element of length of the oriented boundary R of T; if finally the oriented angles of R are denoted by α_k, then (33) *holds.*

It may be remarked that the existence of a field $\boldsymbol{w}$ of constant length and of class C_2 may be proved on T, whenever the boundary R of T is not the empty set. For such a field the considerations of this section prove the formula $\Delta\omega = 2\pi C$, where ω is the angle from $\boldsymbol{w}$ to the tangent vector of R, and $\Delta\omega$ is the variation of ω along R. This formula may be considered as the generalisation of the "Umlaufsatz" for curved surfaces. The proof may of course be based on less regularity assumptions.

10. Stokes Theorem. The result of **7** may be used to give a simple derivation[4] of Stokes' theorem for a vector field and surface with boundary all of class C_1.

Let S, T and R have the properties of **7** in case $n = 1$, and let $\boldsymbol{v}$ be a vector field of class C_1 defined on a region U which has S in its interior.[5] Since Stokes' theorem clearly follows for T if it is proved for the regions $T_{\mathfrak{m}}$ defined in **7**, it may be assumed that S is defined in terms of a single representation (1) and that T is a compact region with a simple closed curve R as boundary.

Let $\boldsymbol{\partial}$ represent a vectorial differential operator of which the action extends only to the vector field $\boldsymbol{v}$. Thus one has, for instance,

$$\boldsymbol{\partial}\cdot\boldsymbol{v} = \operatorname{div}\boldsymbol{v}, \qquad \boldsymbol{\partial}\times\boldsymbol{v} = \operatorname{rot}\boldsymbol{v}, \tag{34}$$

and on the surface S,

$$(\boldsymbol{x}_1\cdot\boldsymbol{\partial})\boldsymbol{v} = \boldsymbol{v}_1, \qquad (\boldsymbol{x}_2\cdot\boldsymbol{\partial})\boldsymbol{v} = \boldsymbol{v}_2. \tag{35}$$

Applying the well-known identity

$$(\boldsymbol{a}\times\boldsymbol{b})\cdot(\boldsymbol{c}\times\boldsymbol{d}) = (\boldsymbol{a}\cdot\boldsymbol{c})(\boldsymbol{b}\cdot\boldsymbol{d}) - (\boldsymbol{a}\cdot\boldsymbol{d})(\boldsymbol{b}\cdot\boldsymbol{c})$$

to $\boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{\partial}, \boldsymbol{v}$ instead of $\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c}, \boldsymbol{d}$, one obtains on S,

$$(\boldsymbol{x}_1\times\boldsymbol{x}_2)\cdot(\boldsymbol{\partial}\times\boldsymbol{v}) = (\boldsymbol{x}_1\cdot\boldsymbol{\partial})(\boldsymbol{v}\cdot\boldsymbol{x}_2) - (\boldsymbol{x}_2\cdot\boldsymbol{\partial})(\boldsymbol{v}\cdot\boldsymbol{x}_1),$$

---

[4] A similar proof is given in McShane's translation of Courant's *Differential und Integral Rechnung.*

[5] It would be sufficient for the boundary R of T to consist of rectifiable curves, since in that case Green may be applied also.

hence by (34) and (35),

$$(\boldsymbol{x}_1\boldsymbol{x}_2 \operatorname{rot} \boldsymbol{v}) = \boldsymbol{v}_1 \cdot \boldsymbol{x}_2 - \boldsymbol{v}_2 \cdot \boldsymbol{x}_1,$$

and by (3),

(36) $$(\operatorname{rot} \boldsymbol{v} \cdot \boldsymbol{x}_3)(\boldsymbol{x}_1\boldsymbol{x}_2\boldsymbol{x}_3) = \boldsymbol{v}_1 \cdot \boldsymbol{x}_2 - \boldsymbol{v}_2 \cdot \boldsymbol{x}_1.$$

The validity of this symbolic computation is obvious.

On applying Green's theorem in the form (29) to the different components of $\boldsymbol{v}$ and $\dot{\boldsymbol{x}}$ one finds

$$\int_R \boldsymbol{v} \cdot \dot{\boldsymbol{x}}\, ds = \int\!\!\int_T (\boldsymbol{v}_1 \cdot \boldsymbol{x}_2 - \boldsymbol{v}_2 \cdot \boldsymbol{x}_1)\, du^1 du^2$$

or, by (36) and (4)

(37) $$\int_R \boldsymbol{v} \cdot \dot{\boldsymbol{x}}\, ds = \int\!\!\int_T \operatorname{rot} \boldsymbol{v} \cdot \boldsymbol{x}_3\, d\sigma.$$

On defining $\dot{\boldsymbol{x}}\, ds = d\boldsymbol{s}$ and $\boldsymbol{x}_3 d\sigma = d\boldsymbol{\sigma}$ to be the vectorial elements of length and area of R and T one obtains the

THEOREM OF STOKES. *If T is a compact region of a surface S of class C_1 and the boundary of T consists of a finite number of arcs of class C_1, if furthermore $d\boldsymbol{s}$ and $d\boldsymbol{\sigma}$ are the vectorial elements of length and area of R and T and $\boldsymbol{v}$ is a vector field of class C_1 in a region U which contains S, then*

(38) $$\int_R \boldsymbol{v} \cdot d\boldsymbol{s} = \int\!\!\int_T \operatorname{rot} \boldsymbol{v} \cdot d\boldsymbol{\sigma}.$$

THE JOHNS HOPKINS UNIVERSITY.

HOMOMORPHISM OF RINGS AND FIELDS OF POINT SETS.*

By Morris Kline.

1. Introduction. Though systems of sets of points such as covering systems, systems of sets defining a space, Borel, and analytic sets have been extensively studied, little attention has been paid to questions involving properties analogous to group properties of such or more general systems. Recently [1] M. H. Stone has studied the abstract algebraic properties of Boolean rings and their connection with Boolean algebras, and his work has significance for systems of point sets which are interpretations of Boolean algebras. The rings and fields of point sets studied in this paper do constitute interpretations of generalized Boolean algebras, and the complete fields, a term to be defined shortly, here studied, constitute an interpretation of the Boolean algebras. Hence some of the group properties of rings and fields of point sets are already known. Yet, because the point set systems are special cases, we are permitted to obtain results for them thus far not obtained for the more general systems and perhaps of importance only for point sets.

The work of this paper has bearing on existing material in that many systems of point sets, such as Borel sets and the system of all subsets of a given set, form complete fields; [2] in that a ring of point sets has an immediate interpretation as a ring of functions and conversely; [3] and in that the notion of homomorphism between set systems is useful in topological problems.[4] Also, the possibilities of combining point set and group properties as in the study of topological groups are interesting.

2. Rings and fields of sets. This part will deal with the simpler algebraic properties of rings and fields of point sets. Not all the results in this part are new; in a few instances, as indicated below, the conclusions are implied by work on Boolean rings. Nevertheless, a few proofs are given of

---

* Received February 10, 1937; revised August 31, 1937.

[1] M. H. Stone, "The theory of representations for Boolean algebras," *Transactions of the American Mathematical Society*, vol. 40 (1936), pp. 37-111.

[2] C. Kuratowski, *Topologie* I, p. 22.

[3] W. Sierpiński, "Sur les anneaux de fonctions," *Fundamenta Mathematicae*, vol. 18 (1932), p. 6.

[4] See, for example, E. Čech, "Théorie générale de l'homologie dans un espace quelconque," *Fundamenta Mathematicae*, vol. 19 (1932), pp. 149-183.

old results because they can be obtained so much more easily for systems of point sets.

DEFINITIONS. *The sum of two systems A and B of point sets, indicated by $A+B$, shall mean the system of sets obtained by adding each set of A to each set of B, as point sets; likewise for the product, indicated by $A \times B$, and the difference, $A - B$, of two systems. A single set is sometimes regarded as a system in the use of these operations.*

A ring of sets is a system of sets to which the sum and intersection of any two sets of the system belong.[5] By a subring we shall understand any subcollection of sets of the ring which also form a ring.

If R is a ring, by the homomorph of R we shall mean a system of sets R' satisfying the following conditions:[6]

1) To each set of R there shall correspond one and only one set of R';

2) Each set of R' shall correspond to at least one set of R;

3) If $A+B=C$ or $A \cdot B = \mathrm{D}$, where A and B are any sets of R, corresponding relations shall hold among the corresponding sets of R'. The relation in R' corresponding to one in R will be the same as the latter unless otherwise specified.

If, instead of condition 2), we have that to each set of R' there corresponds one and only one set of R we shall say that R' is the isomorph of R.

By a residue class of a ring R with respect to a subring S we mean the system $A+S$, where A belongs to R. An ideal is a subring such that if set B belongs to the subring and C is any set of the ring, then $B \cdot C$ belongs to the subring.

The very definition of homomorphism gives the following occasionally useful two theorems.

THEOREM I. *The homomorph of a ring is a ring.*

THEOREM II. *All the residue classes of a ring R with respect to an ideal S form a system of residue classes which is the homomorph of R if we let sum and intersection in R correspond to sum and product, respectively, of the residue classes.*[7]

THEOREM III. *If R is a ring containing a null set,*[8] *the system of residue*

---

[5] F. Hausdorff, *Mengenlehre* (1935), p. 77.

[6] B. L. van der Waerden, *Moderne Algebra*, vol. 1, p. 44.

[7] Cf. van der Waerden, *loc. cit.*, p. 35.

[8] The term null set is here used to mean the empty set or a zero set.

classes determined by any ideal S of R is the isomorph of R under the correspondence of operations of the previous theorem.

Proof. If A and B belong to R, then $A + S = B + S$ if and only if $A = B$. For, since R contains a null set and S is an ideal, S contains a null set. Suppose $A + S = B + S$. Then $A + 0 = B + P$ where P belongs to S. Hence $A \supset B$. Likewise $B + 0 = A + Q$ where Q belongs to S. Hence $B \supset A$, and therefore $A = B$.

It is now obvious that if we let A correspond to $A + S$ the correspondence will, in view of the previous theorem, determine an isomorphism.

DEFINITION. *By a domain of integrity we shall mean a ring in which the intersection of any two non-zero elements is a non-zero element.*

THEOREM IV. *A necessary and sufficient condition that the residue classes with respect to an ideal S of a ring R containing a null set form a domain of integrity*[9] *is that R be a domain of integrity.*

Proof. If A and B belong to R and $A \cdot B = 0$, then the product of the corresponding residue classes would have to give S, else the isomorphism just proven to exist would not hold. Likewise for the converse.

DEFINITION. *If the system R' is the homomorph of the system R, then the sets of R which correspond to the set A' of R' are said to form the class of R corresponding to A'.*[10]

THEOREM V. *If the set system R' is the homomorph of the set system R and R' contains a null set, the class Σ of R corresponding to this null set, $0'$, is an ideal. The other classes are domains of integrity.*[11]

Proof. That Σ is an ideal follows as in van der Waerden. If K_A denotes the system of sets of R corresponding to $A' \neq 0'$ of R', then certainly if A and B belong to K_A, $A + B$ and $A \cdot B$ belong. Moreover, under any homomorphism of two rings, if R contains a null set it must correspond to $0'$ of R'; for, if M be any set of R, since $M \cdot 0 = 0$, $M' \cdot A' = A'$, where 0 corresponds to A'. But M' can be any set of R'; hence A' is included in every set of R' including the null set $0'$. Hence $A' = 0'$. In view of this fact, if A and B belong to K_A, $A \cdot B \neq 0$, else $A' \cdot A' = 0'$. Hence K_A is a domain of integrity.

DEFINITIONS. *A field of sets is a system of sets to which the sum and*

---

[9] The zero residue class is S, itself.

[10] Cf. van der Waerden, *loc. cit.*, p. 32.

[11] Cf. van der Waerden, *loc. cit.*, p. 56.

difference of any two sets of the system belong.[12] *A subfield shall mean any subcollection of sets of the field which is itself a field. By a complete field we shall mean a field to which the sum of all the sets of the field belongs.*

If A and B are any two sets, $A \cdot B = A - [(A + B) - B]$; hence a field is a ring. Hence by the homomorph of a field we shall understand a system of sets satisfying the conditions of a homomorph of a ring. Similarly for isomorphism.

The following result is not new but can be obtained at once for systems of point sets.[13]

THEOREM VI. *Every field of finite order (i. e., field containing a finite number of sets) consists exclusively of sets formed by the process of addition applied to a subcollection of mutually exclusive sets of the field.*

Let the mutually exclusive sets be lettered $B_1, B_2, \cdots, B_n$. Then every set of the field is easily shown to be expressible in the form $A = \sum_{j=1}^{n} b_j B_j$, where each b_j is 0 or 1. From this statement we have

COROLLARY I. *Every field of finite order contains 2^n sets.*[14]

COROLLARY II. *The order of a subfield is a divisor of the order of the field.*

THEOREM VII. *Let the system F' of sets be the homomorph of the field F. Then there exists a field F'' of sets formed from the sets of F', which is the homomorph of F. Moreover, the correspondence between F and F'' holds under the operation of difference.*

Proof. Let the null set of F correspond to the set D' of F'. If A is any set of F, since $A \cdot 0 = 0$, $A' \cdot D' = D'$; and D' is therefore contained in every set of F'.

Let the system F'' consist of all sets of the form $A' - D'$, where A' is any set of F', and let A of F correspond to $A' - D'$ of F''. If the correspondence between F and F'' holds under sum, intersection, and difference it will follow that F'' is a field.

---

[12] Hausdorff, *loc. cit.*, p. 78. Hausdorff uses difference only when the subtrahend is a subset of the minuend. We do not. This distinction is unimportant for fields because, if A and B are any two sets, $A - B = (A + B) - B$ and hence under either use of the term the same sets belong to a field.

[13] Cf. theorems 4 and 12 in the paper of Stone's referred to above.

[14] This result is obtained by B. A. Bernstein, "On finite Boolean algebras," *American Journal of Mathematics*, vol. 57 (1935), p. 742.

That the correspondence holds under sum and intersection is immediate. To show that it holds under difference we have but to employ the fact that $A - B = C$ implies $A = A \cdot B + C$ and $(A \cdot B) \cdot C = 0$.

We should notice that if F' contains a null set, $0'$, then $D' = 0'$ and $F' = F''$. Moreover, if F is a complete field with S as the sum of its sets,[15] then if S' of F' is the correspondent of S, S' must be the sum of the sets of F'. This follows because if A is any set of F we have $S \cdot A = A$; then $S' \cdot A' = A'$. These facts give the

COROLLARY. *If the set system F' is the homomorph of the field (complete field) F, where F' contains a null set, then F' is a field (complete field). Moreover, the correspondence between F and F' holds under the operation of difference.*[16]

It is easy to prove that

THEOREM VIII. *The homomorph F' of a complete field F, under a correspondence of sum and intersection in F with intersection and sum, respectively, in F', is a complete field provided F' contains a null set.*

THEOREM IX. *If F' is a set system which is the homomorph of the field F and if F' contains a null set, then the class Σ of F which corresponds to $0'$ is a field, and the sets of any other class are expressible in terms of the sum and difference of the sets of Σ and some one member of that class.*[17]

Proof. If A and B of F correspond to $0'$ of F', then $A + B$ and $A - B$ correspond to $0'$ because theorem VII assures us that the correspondence holds under difference.

Let A of F correspond to $A' \neq 0'$ of F'. Suppose B of F corresponds to A'. We shall show that B is expressible as the sum and difference of A and the sets of Σ. Let $A - B = X$. Since $A - B$ corresponds to $A' - A' = 0'$, $A - B$ belongs to Σ. Likewise $B - A = Y$ belongs to Σ. Finally, $B = A - X + Y$.

COROLLARY. *If Σ consists of the null set only, F is isomorphic to F'.*

THEOREM X. *If the set system F' is the homomorph of the field F; if F' contains a null set; and if the classes of F each contain a finite number of sets, then any class of F is a residue class with respect to Σ where Σ is the class of F which corresponds to $0'$.*

---

[15] We refer to S later as the unit element of F.

[16] This last statement is contained in theorem 42 of the paper of Stone's referred to above.

[17] Cf. van der Waerden, *loc. cit.*, p. 56.

Proof. Suppose K_A is the class of sets which correspond to A' of F'. Since K_A contains a finite number of sets, the intersection of all these sets, $\bar{A}$ say, is a set of K_A since K_A is a ring by theorem V. As in the preceding theorem, if A belongs to K_A, then $A - \bar{A}$ belongs to Σ. Then the class K_A is the residue class $\bar{A} + \Sigma$, for $A = \bar{A} + (A - \bar{A})$ and this last is a set of $\bar{A} + \Sigma$.

Suppose the points of space T' are the transforms under the single-valued function, f, of the points of space T. If A is a set in T, let us understand the corresponding set in T' to mean the set of points of T' which correspond to the points of A. We then have the

THEOREM XI. *Let the points of space T' be single-valued transforms of the points of space T. Let F be a field, or complete field, of sets in T such that, if A and B are any sets of F for which $A \cdot B = 0$ then for the corresponding sets A' and B' in T', $A' \cdot B' = 0$.* The corresponding sets in T' form a field, or complete field, isomorphic to F.

Proof. Under any single-valued transformation we have that if $A + B = C$, where A and B are here sets of F, then $A' + B' = C'$.

Suppose $A \cdot B = D$. Again we have at once that $D' \subset A' \cdot B'$ and we show first that $D' \supset A' \cdot B'$. Suppose p' belongs to A' and to B'. Then some correspondent of p', p say, is in A, and some correspondent, q say, is in B. If $p = q$, p' is in D'. If $p \neq q$ but both p and q belong to A and B, then p' is in D'. If $p \neq q$, and q does not belong to A say, then we shall have $A \cdot (B - A) = 0$, but $A' \cdot (B - A)' \neq 0$, contrary to hypothesis. Hence $D' = A' \cdot B'$.

We have, so far, that the system F' of sets in T' is the homomorph of F. The corollary to theorem VII gives us the fact that F' is a field (complete if F is complete). We note further that the only set of F which can correspond to the null set of F' is the null set of F. Hence, by the corollary to theorem IX, F' is the isomorph of F.

3. The correspondence of sums and limits of sequences of sets.

It is possible to have two fields isomorphic to each other and for the first field to contain the sum of a countable number of its sets without the second field containing the sum of the corresponding sets. The following is an example of such a situation. Let the field F consist of a system of concentric circles of radii $1/2, 3/4, 7/8, \cdots$, a circle of radius 2, and the difference of any two of these sets. Let the field F' consist of another system of concentric circles of radii $1/2, 3/4, 7/8, \cdots$, the sum of these circles, and the difference of any two of these sets. If we let circles of equal radii correspond and let the circle of radius 2 in the first system correspond to the sum of the sets in the second, the two systems will be isomorphic but the first does not contain the sum of the countable number of circles of radii less than one. Needless to add, the limit sets[18] of topologically convergent sequences of sets of the systems may not belong to either system, much less correspond even when the sets of the sequences do.

Hence we may consider the following questions. Given two isomorphic fields, under what conditions does the sum or intersection of a countable sequence of sets belong to the second field if the sum or intersection of the corresponding sets belongs to the first? Under what conditions do these sums or intersections correspond when they belong? Under what conditions do the limit sets of convergent sequences of corresponding sets belong to the fields and correspond? These questions are considered in this part; part of the first question and others are more conveniently reserved for the next one.

Essentially the problems which follow are concerned with conditions under which the isomorphism between two fields can be extended to new elements of the fields. The theorems hold in separable, metric spaces unless otherwise indicated.

DEFINITIONS. *A σ-ring is a ring to which the sum of each (countable) sequence of sets of the ring belongs. Similarly for σ-fields.*[19] *By a complete σ-field we shall mean a σ-field with unit element.*

THEOREM XII. *Let the complete σ-field F' be the isomorph of the complete σ-field F. Let $\{A_n\}$ be any sequence of sets of F and $\{A'_n\}$ the sequence of corresponding sets of F'. Then $\sum_{n=1}^{\infty} A_n$ corresponds to $\sum_{n=1}^{\infty} A'_n$; $\prod_{n=1}^{\infty} A_n$ and $\prod_{n=1}^{\infty} A'_n$ belong to F and F' respectively and correspond.*

Proof. Let $A = \sum_{n=1}^{\infty} A_n$ of F correspond to A' of F'. Also, let $\sum_{n=1}^{\infty} A'_n$ of F' correspond to B of F. Since $A \cdot A_n = A_n$, $A' \cdot A'_n = A'_n$ and therefore $A' \supset \sum_{n=1}^{\infty} A'_n$. Likewise, $B \supset \sum_{n=1}^{\infty} A_n$. However, this last implies that $\sum_{n=1}^{\infty} A'_n \supset A'$. Hence $A' = \sum_{n=1}^{\infty} A'_n$ and $B = \sum_{n=1}^{\infty} A_n$. This proves the first part of the theorem.

If A_n of F corresponds to A'_n of F', then $C(A_n)$ corresponds to $C(A'_n)$

---

[18] The term limit set will be used to mean the limit set of a sequence of point sets converging in the topological sense. See Kuratowski, *loc. cit.*, p. 155.

[19] F. Hausdorff, *Grundzüge der Mengenlehre*, p. 23.

by the corollary to theorem VII. Then $\sum_{n=1}^{\infty} C(A_n)$ and $\sum_{n=1}^{\infty} C(A'_n)$ belong to F and F' respectively and they correspond to each other by the first part of this theorem. Since $\prod_{n=1}^{\infty} A_n = C(\sum_{n=1}^{\infty} C(A_n))$ and similarly for $\prod_{n=1}^{\infty} A'_n$, the rest of the theorem follows.

THEOREM XIII. *Let the complete σ-field F' be the isomorph of the complete σ-field F so that closed sets of F correspond to closed sets of F' and conversely. If the limit sets of monotonically increasing sequences of sets of F and F' belong to F and F' respectively, then the limit sets of all convergent sequences of sets of F and F' belong to F and F' respectively and correspond when the sets of the sequences do.*

Proof. We first prove that the limit sets of monotonically increasing sequences of corresponding sets correspond. Let $A_1 \subset A_2 \subset A_3 \subset \cdots$ be any monotonically increasing sequence of sets of F. The limit of this sequence is $\overline{\sum_{n=1}^{\infty} A_n}$.[20] The sequence of corresponding sets $\{A'_n\}$ is such that $A'_1 \subset A'_2 \subset A'_3 \subset \cdots$ and has for its limit $\overline{\sum_{n=1}^{\infty} A'_n}$. Suppose $\overline{\sum_{n=1}^{\infty} A_n}$ corresponds to A' of F' and $\overline{\sum_{n=1}^{\infty} A'_n}$ corresponds to B of F. Since $\overline{\sum_{n=1}^{\infty} A_n} \supset \sum_{n=1}^{\infty} A'_n$, $A' \supset \sum_{n=1}^{\infty} A'_n$; and, since by hypothesis A' is closed, $A' \supset \overline{\sum_{n=1}^{\infty} A'_n}$. Likewise $B \supset \overline{\sum_{n=1}^{\infty} A_n}$. Then, by the isomorphism, $\overline{\sum_{n=1}^{\infty} A'_n} \supset A'$. Hence $A' = \overline{\sum_{n=1}^{\infty} A'_n}$ and $B = \overline{\sum_{n=1}^{\infty} A_n}$.

We now note that if A is any set of F and A' the corresponding set of F', then $\bar{A}$, the closure of A, belongs to F, and corresponds to $\bar{A}'$ which belongs to F'. For, if we form the sequence for which $A_n = A$, then $\lim_{n=\infty} A_n = \bar{A}$. Likewise for $\bar{A}'$ and the first part of the proof gives the fact that $\bar{A}$ corresponds to $\bar{A}'$.

Now let $\{A_n\}$ be a sequence of monotonically decreasing sets of F and let $\{A'_n\}$ be the sequence of corresponding sets of F'. Then $\lim_{n=\infty} A_n$ belongs to F; $\lim_{n=\infty} A'_n$ belongs to F'; and the first limit corresponds to the second. For, $\lim_{n=\infty} A_n = \prod_{n=1}^{\infty} \bar{A}_n$[21] and $\lim_{n=\infty} A'_n = \prod_{n=1}^{\infty} \bar{A}'_n$. By the preceding paragraph, $\bar{A}_n$

---

[20] Kuratowski, *loc. cit.*, p. 155.

and A'_n belong to F and F' respectively and correspond. The preceding theorem gives the conclusions asserted in this paragraph.

To complete the proof we utilize the following elementary lemma which is stated without proof.

LEMMA. *If a sequence of sets has a limit, that limit is the limit of a monotonically decreasing sequence of sets formed from the sets of the first sequence by summing.*

To conclude the proof of the theorem we have but to note that if $\{A_n\}$ is any sequence of sets of F and $\{A'_n\}$ the sequence of corresponding sets of F', then the set $B_n = \sum_{k=n}^{\infty} A_k$ corresponds to $B'_n = \sum_{k=n}^{\infty} A'_k$ by theorem XII. By the lemma, $\lim_{n=\infty} A_n = \lim_{n=\infty} B_n$ and likewise for A'_n and B'_n. But $\lim_{n=\infty} B_n$ belongs to F and $\lim_{n=\infty} B'_n$ belongs to F' and the two limits correspond by the results of one of the paragraphs above.

The preceding theorem uses an hypothesis on the correspondence of closed sets which is, in a sense, a generalization of homeomorphism for two spaces. This is apparent if we specialize F and F' to the case where each consists of all the subsets of a given space. The question of for what isomorphic fields this condition on the sets implies a homeomorphism between the unit elements of the fields remains open as does the question of what additional conditions are necessary to imply this conclusion for any two isomorphic, complete σ-fields.

4. The correspondence of open and closed sets. This section will consider primarily conditions under which open and closed sets in one field correspond to open and closed sets in a homomorphic or isomorphic field.

THEOREM XIV. *Let F' be a complete field which is the homomorph of the complete field F. Suppose that limit sets of monotonically increasing (or decreasing) sequences of corresponding sets of F and F' belong to F and F' respectively and correspond. If A is a set of F open in the unit element S of F, then A' is open in S' of F'.*

Proof. Let $C_n = C(A)$.[22] Then $\{C_n\}$ is a monotonically increasing sequence of sets with $\lim_{n=\infty} C_n = \overline{C(A)}$. By the corollary to theorem VII, $C(A)$ corresponds to $C(A')$ and we have, by the hypothesis on limit sets, that $\overline{C(A)}$ corresponds to $\overline{C(A')}$ because the latter is the limit of the sequence for which $C'_n = C(A')$. Since A is open in S, $C(A)$ is closed in S. Hence

---

[21] *Ibid.*

[22] It is to be remembered in this theorem that complements are with respect to the unit elements of the complete fields.

$C(A) = \overline{C(A)}$. Then $C(A') = \overline{C(A')}$, and $C(A')$ is closed in S' or A' is open in S'.

THEOREM XV. *Let F' be a field which is the homomorph of the field F. Suppose the limit sets of monotonically decreasing sequences of corresponding sets of F and F' belong to F and F' respectively, and correspond. Then if $\{A_n\}$ is a sequence of open sets of F such that $\sum_{n=1}^{\infty} A_n$ belongs to F, then $\sum_{n=1}^{\infty} A'_n$ belongs to F', and $\sum_{n=1}^{\infty} A_n$ corresponds to $\sum_{n=1}^{\infty} A'_n$.*

Proof. Let $\sum_{n=1}^{\infty} A_n = A$. Suppose A corresponds to A'. Since $A \cdot A_n = A_n$, $A' \cdot A'_n = A'_n$ and $A' \supset \sum_{n=1}^{\infty} A'_n = B'$. We shall show that $A' - B' = 0$.

Let $C'_n = A' - \sum_{i=1}^{n} A'_i$; then $\{C'_n\}$ is a monotonically decreasing sequence of sets of F' whose limit exists and is $\overline{\prod_{n=1}^{\infty}(A' - \sum_{i=1}^{n} A'_i)}$. Call this limit C'. Since $A' - B' = \prod_{n=1}^{\infty}(A' - \sum_{i=1}^{n} A'_i)$, $C' \supset A' - B'$. A sequence in F corresponding to $\{C'_n\}$ is $\{C_n\} = \{A - \sum_{i=1}^{n} A_i\}$, and this sequence has as limit $C = \overline{\prod_{n=1}^{\infty}(A - \sum_{i=1}^{n} A_i)}$. In view of the hypothesis on limit sets C corresponds to C'. But the A_i's are open sets, as are the sets $\sum_{i=1}^{n} A_i$ for each n. Then $\overline{A - \sum_{i=1}^{n} A_i}$ contains no points of $\sum_{i=1}^{n} A_i$ for such points are interior points of $\sum_{i=1}^{n} A_i$ and hence cannot be limit points of the complement in A. Since $A = \sum_{n=1}^{\infty} A_n$, the value of C shows that $C \cdot A = 0$. In view of the homomorphism, $C' \cdot A' = 0'$. Since $C' \supset A' - B'$ it must be that $A' - B' = 0$, and the theorem follows.

COROLLARY. *Under the hypothesis of the theorem and the added condition that F be complete we have that $\sum_{n=1}^{\infty} A'_n$ is open in S', the unit element of F'.*

Proof. By the corollary to theorem VII, F' is complete. Moreover, since $\sum_{n=1}^{\infty} A_n$ is open in the space, it is open in S, the unit element of F. Then, by theorem XIV, $\sum_{n=1}^{\infty} A'_n$ is open in S'.

THEOREM XVI. *Let the complete field F' be the homomorph of the complete field F. Suppose the limit sets of monotonically decreasing sequences of corresponding sets of F and F' belong to F and F' respectively and correspond. If $\{A_n\}$ is a sequence of sets of F closed in S, the unit element of F, and such that $\prod_{n=1}^{\infty} A_n$ belongs to F, then $\prod_{n=1}^{\infty} A'_n$ belongs to F' and is the correspondent of $\prod_{n=1}^{\infty} A_n$. Moreover, $\prod_{n=1}^{\infty} A'_n$, is closed in S'.*

Proof. Since A_n is closed in S, $C(A_n)$ is open in S. Then, since $\prod_{n=1}^{\infty} A_n$ belongs to F, and since $C(\prod_{n=1}^{\infty} A_n) = \sum_{n=1}^{\infty} C(A_n)$, $\sum_{n=1}^{\infty} C(A_n)$ belongs to F. By theorem XV, $\sum_{n=1}^{\infty} C(A'_n)$ belongs to F' and is the correspondent of $\sum_{n=1}^{\infty} C(A_n)$. But $\prod_{n=1}^{\infty} A'_n = C(\sum_{n=1}^{\infty} C(A'_n))$ and complements correspond under homomorphism.

By the corollary to theorem XV, $\sum_{n=1}^{\infty} C(A'_n)$ is open in S', and hence $\prod_{n=1}^{\infty} A'_n$ is closed in S'.

Several of the theorems in part III depended in part upon the hypothesis that closed sets in a field F correspond under isomorphism to closed sets in a field F'. The following theorem gives sufficient conditions for this to be the case.

THEOREM XVII. *Let F' be a field which is the isomorph of the field F. Let corresponding sets of F and F' be of the same dimension,*[23] *and, moreover, be homogeneously dimensional.*[24] *Let A of F be a set for which $\dim B(A)$*[25] *$< \dim A$, and let the same condition hold for the corresponding set A' of F'. If $\bar{A}$ and $\bar{A}'$ belong to F and F' respectively, then $\bar{A}$ corresponds to $\bar{A}'$.*

Proof. Let $\bar{A}$ of F correspond to B' of F' and let $\bar{A}'$ of F' correspond to C of F. Since $\bar{A}' \supset A'$, $C \supset A$. Likewise $B' \supset A'$. Let $C - \bar{A} = E_1$. Then $\bar{A}' - B' = E'_1$ where E'_1 is the correspondent of E_1. Suppose the dimension of A is k. Since $B(A)$ is a closed set, and hence both an F_σ and a G_δ,[26] and since $\bar{A} = A + B(A)$, $\bar{A}$ is k-dimensional.[27] Then, by hypothesis, B' is k-dimensional. Since A is k-dimensional, again by hypothesis, A' is. Then, by reasoning similar to that just used, $\bar{A}'$ and C are k-dimensional.

---

[23] In the sense of Menger. See his *Dimensionstheorie,* chap. II.

[24] A set is homogeneously dimensional if it is of the same dimension in each of its points.

[25] Menger defines the boundary of a set only for open sets. See his *Dimensionstheorie,* p. 34. We use the more general definition of Kuratowski, *loc. cit.,* p. 24. There the boundary of a set A is defined to be $\overline{A} \cdot \overline{C(A)}$.

Since $B' \supset A'$, E'_1, which is $\bar{A}' - B'$, is a subset of the boundary of A', and, since $\dim B(A') < \dim A'$, $\dim E'_1 < k$.[28] Hence, E_1 is of dimension $< k$. Since $C = C \cdot \bar{A} + E_1$, and since $\bar{A} \cdot E_1 = 0$, C must be at most $(k - 1)$-dimensional in the points of E_1 because such points can be enclosed in neighborhoods whose boundaries do not intersect $\bar{A}$ and hence intersect C in points of E_1 only. However, since C is homogeneously dimensional, $E_1 = 0$. In view of the isomorphism, $E'_1 = 0$. Then $B' \supset \bar{A}'$ and $\bar{A} \supset C$. Now let $\bar{A} - C = E_2$. Then $B' - \bar{A}' = E'_2$ where E'_2 corresponds to E_2. In view of the symmetry due to the isomorphism we may use the process just employed to show that $E'_2 = 0$. Then we have that $B' \subset \bar{A}'$, and therefore that $B' = \bar{A}'$. Hence $\bar{A}$ corresponds to $\bar{A}'$.

The above theorem has application as the following two corollaries show.

COROLLARY I. *Let the hypothesis of the above theorem relating to F and F' hold for F and F' in the Cartesian n-dimensional space, R_n. Let A and A' be corresponding, open sets of F and F' respectively. If $\bar{A}$ belongs to F and $\bar{A}'$ to F', then $\bar{A}$ corresponds to $\bar{A}'$.*

Proof. We have but to apply the theorem[29] that in R_n the boundaries of A and A' are of dimension $\leqq n - 1$.

COROLLARY II. *Let the hypothesis of the above theorem relating to F and F' hold in R_n, where F and F' are now σ-fields. Let $\{A_n\}$ and $\{A'_n\}$ be sequences of corresponding, open sets of F and F' respectively. If F and F' each contain the closures of their sets, then $\overline{\sum_{n=1}^{\infty} A_n}$ and $\overline{\sum_{n=1}^{\infty} A'_n}$ belong to F and F' respectively and correspond.*

Proof. That $\sum_{n=1}^{\infty} A_n$ and $\sum_{n=1}^{\infty} A'_n$ belong to F and F' respectively follows from the definition of σ-fields. They correspond to each other according to theorem XII. Then $\overline{\sum_{n=1}^{\infty} A_n}$ and $\overline{\sum_{n=1}^{\infty} A'_n}$ belong to F and F' respectively by hypothesis and they correspond to each other by the first corollary.

NEW YORK UNIVERSITY.

---

[26] Hausdorff, *Grundzüge der Mengenlehre*, p. 306.

[27] Menger, *loc. cit.*, p. 114.

[28] *Ibid.*, p. 81.

[29] Menger, *loc. cit.*, p. 268.

POLYNOMIAL IDEALS DEFINED BY INFINITELY NEAR BASE POINTS.*

By Oscar Zariski.

Introduction. The linear systems of curves, in the plane or on an algebraic surface, which are theoretically of importance, are the *complete systems.* For complete linear systems the defining linear conditions are *base conditions,* by which the curves of the system are constrained to pass with assigned multiplicities through an assigned set of base points. The set of base points may consist in part of proper points and in part of points infinitely near and in the successive neighborhoods of the proper points ([10], p. 27). However vague this geometric terminology may sound, it is nevertheless true that the facts involved have a precise algebraic meaning, and satisfactory definitions are available in terms of analytical branches and of intersection multiplicities of such branches. But it is equally true that although a well rounded geometric theory can and has been developed along these lines ([1], pp. 327-399), the arithmetic content of the notion of infinitely near points still remains somewhat obscure. It is the main purpose of the present investigation to develop an arithmetic theory parallel to the geometric theory of infinitely near points (in the plane or on a surface without singularities). By this we mean primarily a systematic study of those polynomials ideals in $\mathfrak{k}[x, y]$, or formal power series ideals (in two indeterminates), which adequately describe linear conditions having the character of base conditions. We call these ideals *complete ideals* (II, 12) by analogy with the terminology used in the theory of linear systems. We always suppose that the underlying field $\mathfrak{k}$ is algebraically closed and of characteristic zero, but the theory could be extended with a few modifications to fields of any characteristic. At any rate, the hypothesis that the characteristic is zero is not used in the first four sections of Part I. For possible generalizations to spaces of higher dimension than 2 it would be important to consider also fields which are not algebraically closed.

The class of complete ideals enjoys several striking properties, which respond, however, to a high geometric expectation. *This class is closed under all standard operations on ideals* (except addition) : *the intersection, the product and the quotient of two complete ideals is a complete ideal* (II, 12). Moreover, *a complete ideal has a unique factorization into simple complete ideals* (I, 7), an ideal being simple if it is not the product of ideals different from the unit ideal.

---

* Received November 8, 1937.

We define complete ideals in terms of *valuation ideals.* By valuation ideals in the polynomial ring $\mathfrak{k}[x, y]$ we mean the contracted ideals of the ideals of any valuation ring (belonging to some valuation of the field $\mathfrak{k}(x, y)$) which contains x and y. Valuation ideals are complete ideals, and, in particular, the simple complete ideals are valuation ideals. Most of our work is a study of valuation ideals (briefly: v-ideals) whether from an axiomatic point of view (Part I) or from the point of view of formal power series (Part II). The study of the behaviour of valuation ideals under quadratic transformations (I, 4 and 5) leads to reduction theorems (Theorems 4. 4 and 5. 3) which form the basis of many inductive proofs. *A simple v-ideal of kind* $k+1$ (I, 6) *represents the arithmetic analogue of the notion of a point infinitely near and in the k-th neighborhood of a proper point.*

The treatment deals explicitly only with polynomial rings and rings of holomorphic functions. It is clear, however, that the results carry over automatically to algebraic surfaces without singularities, since any set of base conditions at a simple point P of an algebraic surface is described by a complete ideal in the ring of holomorphic functions of the uniformizing parameters at P.

We make one more remark. In many instances the proofs do not depend on the fact that we have only two indeterminates. On the other hand, in many points a generalization to any number of variables faces new difficulties. In spaces of higher dimension the base conditions may be of a more complicated type: besides base conditions at an isolated base point, it is possible to have infinitely near base curves, or infinitely near base surfaces, etc. At an isolated base point we may also have such base conditions as are given by infinitesimal base curves (the case of an assigned tangent plane at a base point is the simplest example). The algebro-geometric theory has as yet no firm grasp on these eventualities. A generalization of the present treatment to any number of variables would therefore represent not merely an arithmetization of a known chapter of classical algebraic geometry.

CONTENTS.

PART I.

PAGE
1. Valuation ideals in rings of polynomials in two indeterminates. . 153
2. A characteristic property of a Jordan sequence of valuation ideals. 157
3. Further properties of v-ideals. 160
4. v-ideals and quadratic transformations. 163
5. Simple and composite v-ideals. 168
6. Properties of simple v-ideals. 169
7. A factorization theorem for v-ideals. 172

PART II.

PAGE
8. Algebraic and transcendental valuations. 173
9. Valuation ideals in the ring of holomorphic functions. . . 177
10. The notion of a general element of an ideal of formal power series. 179
11. The characterization of simple v-ideals. 184
12. The class of complete ideals. 193
13. Simple v-ideals and divisors of the second kind. . . . 201

PART I.

1. Valuation ideals in rings of polynomials in two indeterminates.

Let $\mathfrak{k}$ be an algebraically closed field of characteristic zero and Σ—a pure transcendental extension of $\mathfrak{k}$ of dimension (degree of transcendentality) two. We consider a valuation B of Σ, i. e. an homorphic mapping of the multiplicative group Σ (the element 0 excluded) upon an ordered abelian group Γ: $a \rightarrow v(a) =$ *value of* a, $a \varepsilon \Sigma$, $a \neq 0$, $v(a) \varepsilon \Gamma$, satisfying the valuation axioms:

$$(1)\ v(a \cdot b) = v(a) + v(b); \qquad (2)\ v(a+b) \geqq \min.(v(a), v(b));$$
$$(3)\ v(a^*) \neq 0,$$

for some a^* in Σ ([2], p. 101; [7]).[1] We assume, moreover, that the elements of the underlying field $\mathfrak{k}$, other than 0, have value zero in the given valuation B.

Let $\mathfrak{B}$ be the valuation ring of B (the set of all elements of Σ whose value is $\geqq 0$). Any ideal $\mathfrak{a}$ in $\mathfrak{B}$ has the following self-evident property: $a \equiv 0(\mathfrak{a})$, $v(b) \geqq v(a)$ implies $b \equiv 0(\mathfrak{a})$. Conversely, any subset of $\mathfrak{B}$ with this property constitutes, together with 0, an ideal. Consequently, given any two ideals $\mathfrak{a}$, $\mathfrak{a}'$ in $\mathfrak{B}$, either $\mathfrak{a} \equiv 0(\mathfrak{a}')$ or $\mathfrak{a}' \equiv 0(\mathfrak{a})$. We say that $\mathfrak{a}$ *precedes* $\mathfrak{a}'$ (and that $\mathfrak{a}'$ *follows* $\mathfrak{a}$) if $\mathfrak{a}' \equiv 0(\mathfrak{a})$, but $\mathfrak{a}' \neq \mathfrak{a}$. The ideals in $\mathfrak{B}$ form then an ordered set. The unit ideal $\mathfrak{B}$ is the first element of this set. The immediate successor of $\mathfrak{B}$ is the ideal $\mathfrak{P}$ consisting of all elements whose value is positive.

Let $\mathfrak{O}$ be a domain of integrity in Σ, *contained in the valuation ring* $\mathfrak{B}$. An ideal $\mathfrak{A}$ in $\mathfrak{O}$ shall be called a *valuation ideal*, or briefly, a *v-ideal, belonging to* or *for the valuation B*, if $\mathfrak{A}$ is the contracted ideal of an ideal $\mathfrak{a}$ in $\mathfrak{B}$, i. e.

---

[1] Note the following consequence of the above axioms: *if* $v(a) < v(b)$, *then* $v(a+b) = v(a)$. Proof: We have $v(a) = v(a \cdot 1) = v(a) + v(1)$, whence $v(1) = 0$. Also $v(-1) + v(-1) = v((-1)^2) = v(1) = 0$, consequently $v(-1) = 0$, since Γ is an ordered group. Hence $v(-b) = v(b)$, and, by axiom (2), $v(a-b) \geqq \min.(v(a), v(b))$. Replacing in this relation a by $a+b$ we find $v(a) \geqq \min.(v(a+b), v(b))$, and our assertion follows.

if $\mathfrak{A} = [\mathfrak{O}, \mathfrak{a}]$. If the reference to the specific valuation B is omitted and if we speak of $\mathfrak{A}$ as a v-ideal, we shall mean then that $\mathfrak{A}$ is a valuation ideal for some valuation B of Σ such that the valuation ring of B contains $\mathfrak{O}$. We have thus defined in $\mathfrak{O}$ a special class of ideals, the class of valuation ideals. We agree to include the zero ideal in this class. Also the valuation ideals in $\mathfrak{O}$, *belonging to a given valuation* B, enjoy the property: $a \equiv 0(\mathfrak{A})$, $v(b) \geqq v(a)$ implies $b \equiv 0(\mathfrak{A})$, where $\mathfrak{A}$ is a v-ideal and a, b are elements of $\mathfrak{O}$. Conversely, any subset of $\mathfrak{O}$ with this property constitutes, together with the zero element of Σ, a valuation ideal belonging to B. The valuation ideals in $\mathfrak{O}$, belonging to a given valuation B, form an ordered set: $\mathfrak{A}$ precedes $\mathfrak{A}'$ if $\mathfrak{A}' \equiv 0(\mathfrak{A})$ and $\mathfrak{A}' \neq \mathfrak{A}$. Since a v-ideal $\mathfrak{A}$ in $\mathfrak{O}$, for the valuation B, is the contracted ideal of an ideal $\mathfrak{a}$ in $\mathfrak{B}$, $\mathfrak{A}$ is also the contracted ideal of its extended ideal $\mathfrak{B}\mathfrak{A}$ in $\mathfrak{B}$. There is thus a $(1, 1)$ correspondence between the v-ideals in $\mathfrak{O}$ belonging to the valuation B and their extended ideals in the valuation ring $\mathfrak{B}$. These extended ideals form in general a proper subset of the set of all ideals of $\mathfrak{B}$.

Let, in particular, $\mathfrak{O}$ be the ring of polynomials in x, y, where we assume that x and y are generating elements of Σ ($\Sigma = k(x, y)$) and elements of the valuation ring.[2] From the fact that every ideal in $\mathfrak{O}$ possesses a finite base and from the valuation axiom (2), it follows that any ideal $\mathfrak{M}$ in $\mathfrak{O}$ contains elements of smallest possible value in B. If α is this minimum, $\alpha = \min\{v(a)\}$, $a \varepsilon \mathfrak{M}$, we shall write $\alpha = v(\mathfrak{M})$ and we shall regard α as *the evaluation of the ideal* $\mathfrak{M}$. Two ideals $\mathfrak{M}$, $\mathfrak{M}'$ will be said to be *equivalent*, in symbols $\mathfrak{M} \sim \mathfrak{M}'$, if $v(\mathfrak{M}) = v(\mathfrak{M}')$. The class $\{\mathfrak{M}\}$ of all ideals equivalent to $\mathfrak{M}$ contains one and only one v-ideal for B, namely the ideal $\mathfrak{A}$ consisting of all elements whose value is $\geqq v(\mathfrak{M})$. Clearly $\mathfrak{A}$ is a divisor of any ideal of the class. In the ordered set of v-ideals of $\mathfrak{O}$, belonging to the valuation B, $\mathfrak{A}$ precedes $\mathfrak{A}'$ if $v(\mathfrak{A}) < v(\mathfrak{A}')$.

In the sequel we shall be dealing with a fixed valuation B and it will be understood that when we speak of a v-ideal in $\mathfrak{O}$ we mean a v-ideal "belonging to the given valuation B."

The valuation ring $\mathfrak{B}$ contains the divisorless ideal $\mathfrak{P}$, consisting of all elements whose value is > 0. We shall denote by $\mathfrak{p}$ the contracted ideal of $\mathfrak{P}$ in $\mathfrak{O}$; $\mathfrak{p}$ is obviously a prime ideal in $\mathfrak{O}$. By the *dimension* r *of the valuation* B is meant the dimension of the ideal $\mathfrak{P}$, i. e. the degree of transcendentality of the field of residual classes $\mathfrak{B}/\mathfrak{P}$ over $\mathfrak{k}$.[3] Evidently, r is either 0 or 1.

---

[2] Note that if x is any element in Σ, then either x or $1/x$ belongs to $\mathfrak{B}$, since $v(1/x) = -v(x)$.

[3] By hypothesis, 0 is the only element of $\mathfrak{k}$ which belongs to $\mathfrak{P}$. Hence $\mathfrak{B}/\mathfrak{P}$ contains a subfield isomorphic to $\mathfrak{k}$. We identify this subfield with $\mathfrak{k}$.

THEOREM 1. *In the ordered set of v-ideals in* $\mathfrak{O}$ *every v-ideal* $\mathfrak{A}$ *has an immediate successor* $\mathfrak{A}'$. *If the valuation B is of dimension zero, then* $\mathfrak{A}'$ *is a maximal subideal of* $\mathfrak{A}$, *and the ring* $\mathfrak{A}/\mathfrak{A}'$ *is isomorphic to the underlying field* $\mathfrak{k}$.

Proof. By the valuation axiom (3), there exists an element a^* such that $v(a^*) > 0$. Let $a^* = f(x,y)/g(x,y)$, $f \varepsilon \mathfrak{O}$, $g \varepsilon \mathfrak{O}$. Since $v(f) = v(a^*) + v(g)$ and $v(g) \geqq 0$, also $v(f) > 0$. If then $\alpha = v(\mathfrak{A})$, there exist in $\mathfrak{A}$ elements whose value is greater than α, for instance, the elements of $\mathfrak{A}f$. The totality of all polynomials whose value is greater than α constitutes, together with the element 0, a v-ideal $\mathfrak{A}'$, contained in $\mathfrak{A}$, and clearly there exist no v-ideals which follow $\mathfrak{A}$ (proper multiples of $\mathfrak{A}$) and precede $\mathfrak{A}'$ (proper divisors of $\mathfrak{A}'$).

Assume that B is of dimension 0, whence $\mathfrak{B}/\mathfrak{P}$ is an algebraic extension of the underlying field $\mathfrak{k}$. Since $\mathfrak{k}$ is algebraically closed, it follows that $\mathfrak{B}/\mathfrak{P}$ is isomorphic to $\mathfrak{k}$. Let f be an element of $\mathfrak{A}$ of smallest possible value, $v(f) = v(\mathfrak{A})$, and let ϕ be any other element in $\mathfrak{A}$. Since $v(\phi) \geqq v(f)$, we have $v(\phi/f) \geqq 0$, i. e. ϕ/f belongs to $\mathfrak{B}$. Hence $\phi/f \equiv c(\mathfrak{P})$, where c is in $\mathfrak{k}$, $v(\phi/f - c) = v[(\phi - cf)/f] > 0$, i. e. $v(\phi - cf) > v(f)$, and consequently $\phi - cf \equiv 0(\mathfrak{A}')$. This shows that $\mathfrak{A}/\mathfrak{A}' \simeq \mathfrak{k}$ and also that $\mathfrak{A}'$ is a maximal subideal of $\mathfrak{A}$, q. e. d.

If B is of dimension 1, it defines an homomorphism of the field Σ upon the field $\mathfrak{B}/\mathfrak{P}$ of algebraic functions of one variable, i. e. B is a "divisor" of Σ. If $\mathfrak{O}$ contains elements which *mod* $\mathfrak{p}$ are transcendental with respect to $\mathfrak{k}$, i. e. if $\mathfrak{p}$ is a 1-dimensional ideal in $\mathfrak{O}$, we are dealing with a *divisor of the first kind* with respect to $\mathfrak{O}$. The ideal $\mathfrak{p}$ is then a principal ideal, say $\mathfrak{p} = (f)$, where f is an irreducible polynomial, and the v-ideals in $\mathfrak{O}$ are the ideals $\mathfrak{p}^n = (f^n)$, $n = 0, 1, 2, \cdots$. If, however, $\mathfrak{p}$ is 0-dimensional, we are dealing with a *divisor of the second kind* with respect to $\mathfrak{O}$. The v-ideals in $\mathfrak{O}$ are in this case certain primary ideals belonging to $\mathfrak{p}$. We need not consider separately this case, because it reduces to the case of 0-dimensional valuations. In fact, we may consider an arbitrary valuation B_1 of the field $\Sigma_1 = \mathfrak{B}/\mathfrak{P}$ (a point of the Riemann surface of the field Σ_1). The given valuation B of Σ followed up by the valuation B_1 of Σ_1 defines an homomorphism of Σ upon the underlying field $\mathfrak{k}$ (together with symbol ∞), hence a 0-dimensional valuation B' of Σ. The v-ideals in $\mathfrak{O}$ belonging to B will be among the v-ideals belonging to B'.

From now on we shall only consider 0-dimensional valuations. If B is 0-dimensional, then the v-ideal $\mathfrak{p} = [\mathfrak{P}, \mathfrak{O}]$ is prime and 0-dimensional, since it is the immediate successor of the unit ideal $\mathfrak{O}$ and since, by Theorem 1, $\mathfrak{O}/\mathfrak{p} \simeq \mathfrak{k}$. Replacing, if necessary, x and y by $x - c$, $y - d$, where $x \equiv c(\mathfrak{p})$, $y \equiv d(\mathfrak{p})$, we may assume that $x \equiv 0(\mathfrak{p})$, $y \equiv 0(\mathfrak{p})$, whence $\mathfrak{p} = (x, y)$.

Starting with $\mathfrak{O} = \mathfrak{q}_0$ and with its successor $\mathfrak{p} = \mathfrak{q}_1$, we form the simple sequence of v-ideals

$$(1) \qquad \mathfrak{q}_0, \mathfrak{q}_1, \mathfrak{q}_2, \cdots, \mathfrak{q}_i, \cdots$$

where $\mathfrak{q}_{i+1}$ is the immediate successor of $\mathfrak{q}_i$. *Each $\mathfrak{q}_{i+1}$ is a maximal subideal of its predecessor $\mathfrak{q}_i$.* We shall call a *Jordan sequence* any sequence of ideals having this last mentioned property. It is clear that all the ideals $\mathfrak{q}_i$, $i \geqq 1$, in the Jordan sequence (1), are primary ideals belonging to $\mathfrak{p}$ $(= \mathfrak{q}_1)$. In fact: (1) $\mathfrak{q}_i \equiv 0(\mathfrak{p})$; (2) $ab \equiv 0(\mathfrak{q})$, $a \not\equiv 0(\mathfrak{q})$ imply that $v(a) + v(b) > v(a)$, whence $v(b) > 0$ and $b \equiv 0(\mathfrak{p})$; (3) $v(\mathfrak{q}_1) = v(\mathfrak{p}) < v(\mathfrak{p}^2) < v(\mathfrak{p}^3) < \cdots < v(\mathfrak{p}^i)$, whence $v(\mathfrak{p}^i) \geqq v(\mathfrak{q}_i)$ and consequently $\mathfrak{p}^i \equiv 0(\mathfrak{q}_i)$.

Two cases are possible: (a) either the intersection of all the ideals $\mathfrak{q}_i$ is the zero-ideal; (b) or this intersection is a certain ideal $\mathfrak{p}_1 \neq (0)$. In the first case the Jordan sequence (1) contains *all* the v-ideals of $\mathfrak{O}$ belonging to the given valuation B. We investigate now the second case.

We first prove that *$\mathfrak{p}_1$ is a prime ideal.* In fact, let $ab \equiv 0(\mathfrak{p}_1)$, $a \not\equiv 0(\mathfrak{p}_1)$, and let $\mathfrak{q}_r$ be the last ideal in the sequence $\{\mathfrak{q}_i\}$ which contains a. Then $v(a) = v(\mathfrak{q}_r)$, whence $v(a) \leqq \rho v(\mathfrak{p})$, where ρ is the exponent of the ideal $\mathfrak{q}_r$. No power $\mathfrak{p}^n$ of $\mathfrak{p}$ can belong to all the ideals $\mathfrak{q}_i$, since $\mathfrak{p}^n/\mathfrak{p}$ is of finite rank with respect to $\mathfrak{k}$. Hence, for any integer $n > 0$ there exists an integer α_n such that $v(\mathfrak{q}_{\alpha_n}) > v(\mathfrak{p}^n)$. Since $ab \equiv 0(\mathfrak{q}_i)$ for any value of i, it follows that $v(a) + v(b) > nv(\mathfrak{p})$, n arbitrary. In view of the inequality $v(a) \leqq \rho v(\mathfrak{p})$, we deduce that also $v(b) > nv(\mathfrak{p})$, n—an arbitrary integer. In particular, if $\mathfrak{q}_m$ is any ideal in the sequence $\{\mathfrak{q}_i\}$ and if ρ_m is its exponent, we will have $v(\mathfrak{q}_m) \leqq \rho_m v(\mathfrak{p}) < v(b)$, whence $b \equiv 0(\mathfrak{q}_m)$, for any m. It follows that $b \equiv 0(\mathfrak{p}_1)$, i. e. $\mathfrak{p}_1$ is a prime ideal.

The ideal $\mathfrak{p}_1$ is one-dimensional, say (f), where f is an irreducible polynomial. The inequality $v(b) > nv(\mathfrak{p})$, n arbitrary, holds for any element b of $\mathfrak{p}_1$, and this shows that the value group of B is non-archimedean (B is a "special" valuation, of rank 2. See [2], p. 113). It is not difficult to see that *all the v-ideals in $\mathfrak{O}$ for B are of the form $\mathfrak{p}_1^m\mathfrak{q}_n$, $m, n = 0, 1, 2 \cdots$.* In fact, let F_1 and F_2 be any two polynomials in $\mathfrak{O}$ and let $F_1 = f^{m_1}G_1$, $F_2 = f^{m_2}G_2$, where G_1 and G_2 are not divisible by f. If $m_1 > m_2$, then $v(F_1) > v(F_2)$, since $v(f) > v(G_2)$. If $m_1 = m_2$, then $v(F_1) > v(F_2)$ or $v(F_1) = v(F_2)$ according as $v(G_1) > v(G_2)$ or $v(G_1) = v(G_2)$. Hence the set of all polynomials whose value is not less than the value of a given polynomial f^mG

$(G \not\equiv 0(f))$ coincides with the ideal $\mathfrak{p}_1^m\mathfrak{q}_n$, where $v(\mathfrak{q}_n) = v(G)$. This form of the v-ideals brings out clearly the well-known decomposition of the valuation B into two valuations of rank 1 and the nature of the value group, as consisting in this case of pairs of integers. B decomposes into two valuations, B' and $\bar{B}$. B' is the one-dimensional valuation defined by the prime 1-dimensional ideal $\mathfrak{p}_1$, and its valuation ring $\mathfrak{B}'$ is the set of rational functions $\dfrac{F(x,y)}{G(x,y)}$, $G \not\equiv 0(f)$. B' maps Σ upon the field $\bar{\Sigma} = \mathfrak{k}(\bar{x}, \bar{y})$ of algebraic functions of one variable, where $\bar{\Sigma}$ is the quotient field of the ring $\mathfrak{O}/\mathfrak{p}_1$. $\bar{B}$ is a valuation of $\bar{\Sigma}$, and the sequence $\{\mathfrak{q}_i\}$ (with x, y replaced by $\bar{x}, \bar{y}$) is the sequence of the valuation ideals in the ring $\bar{\mathfrak{O}} = \mathfrak{k}[\bar{x}, \bar{y}]$ which belong to $\bar{B}$.[4]

2. A characteristic property of Jordan sequence of valuation ideals. Given a Jordan sequence of ideals in $\mathfrak{O}$: $\mathfrak{q}_0, \mathfrak{q}_1, \mathfrak{q}_2, \cdots$, where $\mathfrak{q}_0 = \mathfrak{O}$, $\mathfrak{q}_1 = \mathfrak{p} = (x, y)$, we ask under what conditions will there exist a valuation of Σ for which the given sequence $\{\mathfrak{q}_i\}$ is the sequence of (zero-dimensional) v-ideals. In other words: *under what conditions does the given sequence* $\{\mathfrak{q}_i\}$ *belong to a valuation of* Σ*?*

THEOREM 2. 1. *A necessary and sufficient condition in order that a Jordan sequence* $\{\mathfrak{q}_i\}$ *of* 0*-dimensional ideals in* $\mathfrak{O}$ *belong to a valuation of the field* Σ, *is that the quotient* $\mathfrak{q}_i : (a)$ *belong to the sequence, for any* i *and for any elements* a *in* $\mathfrak{O}$.

THEOREM 2. 2. *A necessary and sufficient condition in order that a Jordan sequence* $\{\mathfrak{q}_i\}$ *of* 0*-dimensional ideals in* $\mathfrak{O}$ *belong to a valuation of the field* Σ, *is that the congruences*

$$\mathfrak{q}_i\mathfrak{q}_{j+1} : \mathfrak{q}_j \equiv 0(\mathfrak{q}_{i+1}) \tag{2}$$

hold true for any pair $\mathfrak{q}_i$, $\mathfrak{q}_j$ *of ideals of the sequence.*

The characterization given in Theorem 2. 2 has the advantage of involving only operations within the given sequence $\{\mathfrak{q}_i\}$. We prove both theorems simultaneously.

(1) *The conditions are necessary.* As to Theorem 2. 1, let $v(a) = \alpha$, $\mathfrak{q}_i : (a) = \mathfrak{q}'$, and let b, c be any two elements of $\mathfrak{O}$ such that $b \equiv 0(\mathfrak{q}')$, $v(c) \geqq v(b)$. Since $ba \equiv 0(\mathfrak{q}_i)$, we have $v(c) + v(a) \geqq v(b) + v(a) \geqq v(\mathfrak{q}_i)$, whence $ca \equiv 0(\mathfrak{q}_i)$, and consequently $c \equiv 0(\mathfrak{q}')$. Hence $\mathfrak{q}'$ enjoys the property:

---

[4] It is possible to have $\mathfrak{p}_1 = (0)$ also in the case of a valuation of rank 2. This happens when the component B' (divisor of Σ) of the valuation B, of rank 2, is a divisor of the second kind with respect to $\mathfrak{O}$. (In geometric terms: the divisor B' is an exceptional curve which has been transformed into a point of the plane (x, y)).

$b \equiv 0(\mathfrak{q}')$, $v(c) \geqq v(b)$, implies $c \equiv 0(\mathfrak{q}')$, and is therefore a v-ideal belonging to the given valuation B. Since $\mathfrak{q}_i \equiv 0(\mathfrak{q}')$, necessarily $\mathfrak{q}' = \mathfrak{q}_j$, for some $j \leqq i$.

The necessity of the condition of Theorem 2.2 is proved in a similar manner. Let b be an element of $\mathfrak{q}_i\mathfrak{q}_{j+1} : \mathfrak{q}_j$, whence $b\mathfrak{q}_j \equiv 0(\mathfrak{q}_i\mathfrak{q}_{j+1})$. Since $v(\mathfrak{q}_i\mathfrak{q}_{j+1}) = v(\mathfrak{q}_i) + v(\mathfrak{q}_{j+1})$, we have $v(b) + v(\mathfrak{q}_j) \geqq v(\mathfrak{q}_i) + v(\mathfrak{q}_{j+1})$, and since $v(\mathfrak{q}_j) < v(\mathfrak{q}_{j+1})$, it follows, $v(b) > v(\mathfrak{q}_i)$, i.e. $b \equiv 0(\mathfrak{q}_{i+1})$, q. e. d.

(2) *The conditions are sufficient.* We introduce the following notations: if ξ, η are elements of $\mathfrak{O}$, we write $\xi \leqq \eta$ (or $\eta \geqq \xi$), if the congruence $\xi \equiv 0(\mathfrak{q}_i)$ always implies $\eta \equiv 0(\mathfrak{q}_i)$; we write $\xi < \eta$ (or $\eta > \xi$) if there exists in the sequence $\{\mathfrak{q}_i\}$ an ideal $\mathfrak{q}_m$ such that $\eta \equiv 0(\mathfrak{q}_m)$, $\xi \not\equiv 0(\mathfrak{q}_m)$. We now prove the following lemma, assuming that the condition of Theorem 2.1 *or* that of Theorem 2.2 is satisfied.

LEMMA. *If ξ, η, ζ are elements of $\mathfrak{O}$ and if $\xi\eta \equiv 0(\mathfrak{q}_i)$, then $\eta \leqq \zeta$ implies $\xi\zeta \equiv 0(\mathfrak{q}_i)$ and $\eta < \zeta$ implies $\xi\zeta \equiv 0(\mathfrak{q}_{i+1})$.*

In other words: *$\eta \leqq \zeta$ implies $\xi\eta \leqq \xi\zeta$; $\eta < \zeta$ implies that either $\xi\eta < \xi\zeta$ or that $\xi\eta$ and $\xi\zeta$ belong to all the ideals $\mathfrak{q}_i$ of the sequence.*

Assume the condition of Theorem 2.1, and let $\mathfrak{q}_i : (\xi) = \mathfrak{q}_h$. Since $\eta \equiv 0(\mathfrak{q}_h)$, it follows that if $\eta \leqq \zeta$, then also $\zeta \equiv 0(\mathfrak{q}_h)$, whence $\xi\zeta \equiv 0(\mathfrak{q}_i)$, and this proves the first part of the lemma. Let $\mathfrak{q}_{i+1} : (\xi) = \mathfrak{q}_s$, $s \geqq h$, and let ϕ and ψ be any pair of elements of $\mathfrak{q}_h$. We have $\phi\xi \equiv 0(\mathfrak{q}_i)$, $\psi\xi \equiv 0(\mathfrak{q}_i)$, whence there exist elements c, d in the underlying field $\mathfrak{k}$ such that $c\phi\xi + d\psi\xi \equiv 0(\mathfrak{q}_{i+1})$, since $\mathfrak{q}_i/\mathfrak{q}_{i+1}$ is of rank 1 with respect to $\mathfrak{k}$. Hence $c\phi + d\psi \equiv 0(\mathfrak{q}_s)$, for any two elements ϕ, ψ in $\mathfrak{q}_h$ and for appropriate elements c, d in $\mathfrak{k}$, i.e. $\mathfrak{q}_h/\mathfrak{q}_s$ is at most of rank 1 with respect to $\mathfrak{k}$, and s is either h or $h+1$. Since $\eta \equiv 0(\mathfrak{q}_h)$, it follows that if $\eta < \zeta$, then $\zeta \equiv 0(\mathfrak{q}_{h+1}) = 0(\mathfrak{q}_s)$, whence $\xi\zeta = 0(\mathfrak{q}_{i+1})$, q. e. d.

Assume the condition of Theorem 2.2. If η belongs to all the ideals $\mathfrak{q}_i$ of the sequence, the same will be true for ζ, and both parts of the lemma are trivial. Assume that there exists a *last* ideal $\mathfrak{q}_h$ which contains $\eta : \eta \equiv 0(\mathfrak{q}_h)$, $\eta \not\equiv 0(\mathfrak{q}_{h+1})$. If $\eta \leqq \zeta$, then also $\zeta \equiv 0(\mathfrak{q}_h)$, whence $\xi\eta\zeta \equiv 0(\mathfrak{q}_i\mathfrak{q}_h)$. Assume, if possible, $\xi\zeta \not\equiv 0(\mathfrak{q}_i)$. There will then exist an ideal $\mathfrak{q}_j$ in our sequence, $j < i$, such that $\xi\zeta \equiv 0(\mathfrak{q}_j)$, $\xi\zeta \not\equiv 0(\mathfrak{q}_{j+1})$. Since $j + 1 \leqq i$, $\mathfrak{q}_i \equiv 0(\mathfrak{q}_{j+1})$, the congruence $\xi\eta\zeta \equiv 0(\mathfrak{q}_i\mathfrak{q}_h)$ implies $\xi\eta\zeta \equiv 0(\mathfrak{q}_{j+1}\mathfrak{q}_h)$. Now $\mathfrak{q}_{j+1}$ is a maximal subideal of $\mathfrak{q}_j$ and $\xi\zeta$ is in $\mathfrak{q}_j$ but not in $\mathfrak{q}_{j+1}$; hence $\mathfrak{q}_j = (\xi\zeta, \mathfrak{q}_{j+1})$, and consequently, $\eta\mathfrak{q}_j = (\eta\xi\zeta, \eta\mathfrak{q}_{j+1}) \equiv 0(\mathfrak{q}_{j+1}\mathfrak{q}_h)$, since $\eta \equiv 0(\mathfrak{q}_h)$. It follows that $\eta \equiv 0(\mathfrak{q}_{j+1}\mathfrak{q}_h : \mathfrak{q}_j) \equiv 0(\mathfrak{q}_{h+1})$, in contradiction with our hypothesis $\eta \not\equiv 0(\mathfrak{q}_{h+1})$. This proves the first part of the lemma.

Let now $\eta < \zeta$, and hence $\zeta \equiv 0(\mathfrak{q}_{h+1})$. We have then $\xi\eta\zeta \equiv 0(\mathfrak{q}_i\mathfrak{q}_{h+1})$

and also by the first part of the lemma, just proved, $\xi\zeta \equiv 0(\mathfrak{q}_i)$. Since $\mathfrak{q}_{h+1}$ is a maximal subideal of $\mathfrak{q}_h$ and since η is in $\mathfrak{q}_h$ but not in $\mathfrak{q}_{h+1}$, we have $\mathfrak{q}_h = (\eta, \mathfrak{q}_{h+1})$, and consequently $\xi\zeta\mathfrak{q}_h = (\xi\zeta\eta, \xi\zeta\mathfrak{q}_{h+1}) \equiv 0(\mathfrak{q}_i\mathfrak{q}_{h+1})$. Hence $\xi\zeta \equiv 0(\mathfrak{q}_i\mathfrak{q}_{h+1} : \mathfrak{q}_h) \equiv 0(\mathfrak{q}_{i+1})$, q. e. d.

The rest of the proof is based solely on the above Lemma. Let $\mathfrak{p}_1$ be the intersection of the ideals $\mathfrak{q}_i$ of our sequence. *We prove that* $\mathfrak{p}_1$ *is a prime ideal.* We first observe, that x and y cannot both belong to $\mathfrak{q}_2$, since $\mathfrak{q}_1 = \mathfrak{p} = (x, y)$. Let, for instance, $x \equiv 0(\mathfrak{q}_1)$, $x \not\equiv 0(\mathfrak{q}_2)$, whence $y \geqq x$. We deduce from our lemma, that if a given power of x, say x^m, belongs to an ideal $\mathfrak{q}_i$ of the sequence, then also all the power products $x^k y^l$, $k + l \geqq m$, belong to $\mathfrak{q}_i$, i. e. $\mathfrak{p}^m \equiv 0(\mathfrak{q}_i)$. Since $\mathfrak{p}^m/\mathfrak{p}$ is of finite rank with respect to $\mathfrak{k}$, it follows that *no power of x can belong to all the ideals* $\mathfrak{q}_i$, *i. e.* to $\mathfrak{p}_1$.

We next observe that *all the ideals* $\mathfrak{q}_i$, $i > 1$, *are primary ideals belonging to* $\mathfrak{p}$ $(= \mathfrak{q}_1)$. In fact: (1) $\mathfrak{q}_i \equiv 0(\mathfrak{p})$. (2) Let $ab \equiv 0(\mathfrak{q}_i)$, $a \not\equiv 0(\mathfrak{q}_i)$, whence $a < ab$. It is not possible to have $b \leqq 1$, because this would imply, by our lemma, $ab \leqq a$, in contradiction with $ab > a$. Hence $1 < b$ and since 1 is in $\mathfrak{q}_0$, b must be in $\mathfrak{q}_1$, i. e. in $\mathfrak{p}$. (3) We have $1 < x$, hence, by our Lemma, $x < x^2$ (since no power of x belongs to $\mathfrak{p}_1$), whence, again by the Lemma, $x^2 < x^3$, and generally, $x < x^2 < x^3 < \cdots < x^i$. Consequently $x^2 \equiv 0(\mathfrak{q}_2)$, $x^3 \equiv 0(\mathfrak{q}_3), \cdots, x^i \equiv 0(\mathfrak{q}_i)$. The congruence $x^i \equiv 0(\mathfrak{q}_i)$ implies, as we have just shown, the congruence $\mathfrak{p}^i \equiv 0(\mathfrak{q}_i)$.

To prove that $\mathfrak{p}_1$ is prime, let $\xi\eta \equiv 0(\mathfrak{p}_1)$ and assume, if possible, that $\xi \not\equiv 0(\mathfrak{p}_1)$, $\eta \not\equiv 0(\mathfrak{p}_1)$. Let $\mathfrak{q}_r$ be the last ideal of the sequence $\{\mathfrak{q}_i\}$ which contains ξ and let similarly $\mathfrak{q}_s$ be the last ideal containing η. If m and n are the exponents of the primary ideals $\mathfrak{q}_r$ and $\mathfrak{q}_s$ respectively, we have $x^m \equiv 0(\mathfrak{q}_r)$, $x^n \equiv 0(\mathfrak{q}_s)$, whence $x^m \geqq \xi$, $x^n \geqq \eta$, and consequently, by our lemma, $x^{m+n} \geqq \xi\eta$, i. e. $x^{m+n} \equiv 0(\mathfrak{p}_1)$, and this is impossible. Hence $\mathfrak{p}_1$ is prime, necessarily either the zero ideal or one-dimensional, since $\mathfrak{p}_1 \equiv 0(\mathfrak{p})$.

We now consider the well ordered descending set S of the ideals $\mathfrak{q}_{mn} = \mathfrak{p}_1{}^m\mathfrak{q}_n$, where clearly $\mathfrak{q}_{mn} \equiv 0(\mathfrak{q}_{m_1n_1})$ if $m > m_1$ or if $m = m_1$ and $n > n_1$. Here $\mathfrak{q}_{0n} = \mathfrak{q}_n$ and if $\mathfrak{p}_1 = (0)$ it is understood that S coincides with the sequence $\{\mathfrak{q}_n\}$. In either case, the intersection of all the ideals $\mathfrak{q}_{mn}$ of the set is the zero ideal. Hence, given any element ξ in $\mathfrak{O}$, there will exist a *first* ideal in the set S, say $\mathfrak{q}_{ij}$, which does not contain ξ. Let η be another element in $\mathfrak{O}$ and let $\mathfrak{q}_{i'j'}$ be the first ideal in the set which does not contain η. We complete and modify our notations $\xi \leqq \eta$, $\xi < \eta$ introduced above, as follows: we write $\xi \gtrdot \eta$ if $\mathfrak{q}_{i'j'} \equiv 0(\mathfrak{q}_{ij})$, and $\xi < \eta$ if $\mathfrak{q}_{i'j'} \equiv 0(\mathfrak{q}_{ij})$ and $\mathfrak{q}_{i'j'} \neq \mathfrak{q}_{ij}$. It is obvious that if $\xi \gtrdot \eta$ and $\eta \gtrdot \zeta$, then $\xi \gtrdot \zeta$, and if $\xi < \eta$ and $\eta \gtrdot \zeta$, or if $\xi \gtrdot \eta$ and $\eta < \zeta$, then $\xi < \zeta$. From our Lemma and from the fact that $\mathfrak{p}_1$ is a principal ideal, it follows in a straight-forward manner that the relations $\xi \not> \eta$, $\xi < \eta$ imply, for any element ζ in $\mathfrak{O}$, the relations $\xi\zeta \not> \eta\zeta$, $\xi\zeta < \eta\zeta$ respectively. It is also evident that if $\xi \not> \eta$, $\xi \not> \eta'$, then $\xi \not> \eta \pm \eta'$, and if $\xi < \eta$ and $\xi < \eta'$ then $\xi < \eta \pm \eta'$.

Let $\mathfrak{B}$ be the set of all elements in the field Σ which can be put in the form $\eta/\xi, \eta, \xi \varepsilon \mathfrak{O}, \xi \not> \eta$. *We prove that* $\mathfrak{B}$ *is a valuation ring.*

First, $\mathfrak{B}$ *is a ring.* In fact, let $\eta/\xi \, \varepsilon \, \mathfrak{B}$, $\eta_1/\xi_1 \, \varepsilon \, \mathfrak{B}$. Since $\xi \not> \eta$ and $\xi_1 \not> \eta_1$, we have $\xi\xi_1 \not> \eta\xi_1$ and $\xi\xi_1 \not> \xi\eta_1$, whence $\xi\xi_1 \not> \eta\xi_1 \pm \xi\eta_1$, i. e.

$$\frac{\eta}{\xi} \pm \frac{\eta_1}{\xi_1} = \frac{\eta\xi_1 \pm \xi\eta_1}{\xi\xi_1} \, \varepsilon \, \mathfrak{B}.$$

Also, since $\xi\xi_1 \not> \xi\eta_1 \not> \eta\eta_1$, the product $\eta\eta_1/\xi\xi_1$ belongs to $\mathfrak{B}$.

To prove that $\mathfrak{B}$ is a valuation ring, it is sufficient to show that given any two elements a, b in $\mathfrak{B}$, then either ab^{-1} or $a^{-1}b$ is in $\mathfrak{B}$ ([2], p. 102). In other words, we have to show that given any two elements ξ, η in $\mathfrak{O}$, either ξ/η or η/ξ must belong to $\mathfrak{B}$. But this is obvious, since one of the two relations $\xi \not> \eta$, $\eta \not> \xi$ must hold true.

Finally, the valuation abstractly defined by the valuation ring $\mathfrak{B}$ is not the trivial one, in which every element has value zero. In other words, the ring $\mathfrak{B}$ does not contain all the elements of the field Σ. In fact, take two elements ξ, η in $\mathfrak{O}$ such that definitely, $\xi < \eta$. We assert that ξ/η does not belong to $\mathfrak{B}$. Assuming the contrary, we must be able to put ξ/η in the form η_1/ξ_1, where $\xi_1 \not> \eta_1$ and $\xi\xi_1 = \eta\eta_1$. By our lemma, $\xi_1 \not> \eta_1$ implies $\eta\xi_1 \not> \eta\eta_1$, while $\xi < \eta$ implies $\xi\xi_1 < \eta\xi_1$, i. e. $\eta\eta_1 < \eta\xi_1$, giving two contradictory relations.

It remains to show that the ideals $\mathfrak{q}_{mn}$ of our well ordered set S are the v-ideals belonging to the valuation B defined by the valuation ring $\mathfrak{B}$. Consider any ideal $\mathfrak{q}_{mn}$ and let ξ, η be elements in $\mathfrak{O}$ such that $\xi \equiv 0(\mathfrak{q}_{mn})$, $v(\eta) \geqq v(\xi)$. Since $v(\eta/\xi) \geqq 0$, we must have $\eta/\xi = \eta_1/\xi_1$, where $\xi_1 \not> \eta_1$. Hence $\xi\xi_1 \not> \xi\eta_1$, i. e. $\xi\xi_1 \not> \eta\xi_1$, and this implies $\xi \not> \eta$. Hence $\eta \equiv 0(\mathfrak{q}_{mn})$, and thus $\mathfrak{q}_{mn}$ enjoys the property: $\xi \equiv 0(\mathfrak{q}_{mn}), v(\eta) \geqq v(\xi)$ implies $\eta \equiv 0(\mathfrak{q}_{mn})$. Hence $\mathfrak{q}_{mn}$ is a v-ideal belonging to the valuation B. That the set $S = \{\mathfrak{q}_{mn}\}$ contains all the v-ideals for B, is implied by the fact that each $\mathfrak{q}_{mn}$ $(n \neq 0)$ is a maximal subideal of its immediate predecessor $\mathfrak{q}_{m,n-1}$, and that $\mathfrak{q}_{m0}$ is the intersection of the ideals which precede it in S.

3. Further properties of v-ideals. We consider the sequence $\{\mathfrak{q}_i\}$ of 0-dimensional v-ideals in $\mathfrak{O}$ belonging to a fixed valuation B of Σ, where $\mathfrak{q}_1 = \mathfrak{p} = (x, y)$ and $\mathfrak{q}_0 = \mathfrak{O}$. We may assume $x \not\equiv 0(\mathfrak{q}_2)$. Since $\mathfrak{q}_1/\mathfrak{q}_2$ is of rank 1 with respect to $\mathfrak{k}$, there exists elements c, d in $\mathfrak{k}$ such that $cx + dy \equiv 0(\mathfrak{q}_2)$, $d \neq 0$. We then replace y by $cx + dy$ and thus we may assume $y \equiv 0(\mathfrak{q}_2)$, whence $\mathfrak{q}_2 = (y, x^2)$.

Let $\mathfrak{p}^h$ be the highest power of $\mathfrak{p}$ which divides a given v-ideal $\mathfrak{q}_i$; $\mathfrak{q}_i \equiv 0(\mathfrak{p}^h)$, $\mathfrak{q}_i \not\equiv 0(\mathfrak{p}^{h+1})$. Every polynomial f in $\mathfrak{q}_i$ is then of the form $f = f_h + f_{h+1}, \cdots$, where f_i is homogeneous of degree i in x and y. We shall call f_h the *subform* of f (for particular polynomials in $\mathfrak{q}_i$, f_h may be identically zero). As f varies in $\mathfrak{q}_i$, its subform f_h generates a linear system of forms, of a certain dimension $r \geqq 0$ (a $\mathfrak{k}$-module of rank $r+1$). Denote this system by $\Omega(\mathfrak{q}_i)$. We define in a similar manner the symbol $\Omega(\mathfrak{A})$ for any ideal $\mathfrak{A}$ in $\mathfrak{O}$.

THEOREM 3. *Let* $\mathfrak{q}_i \equiv 0(\mathfrak{p}^h)$, $\mathfrak{q}_i \not\equiv 0(\mathfrak{p}^{h+1})$ *and let* $\mathfrak{A} = [\mathfrak{q}_i, \mathfrak{p}^k]$, $k \geqq h$. *If* $\Omega(\mathfrak{A})$ *is of dimension* r, *then* $\Omega(\mathfrak{A})$ *coincides with the system of forms (of degree* k*) which are divisible by* y^{k-r}, *and, moreover, there exists a v-ideal* $\mathfrak{q}_j$ *in the sequence* $\{\mathfrak{q}_n\}$ *such that* $\mathfrak{A} = \mathfrak{p}^r\mathfrak{q}_j$.

Proof. Since $\mathfrak{q}_i$ contains polynomials ϕ whose subforms are exactly of degree h, it also contains polynomials (such as $x^{k-h}\phi$) whose subforms are of degree k. Hence $\mathfrak{A} \not\equiv 0(\mathfrak{p}^{k+1})$, and $\Omega(\mathfrak{A})$ consists of forms of degree k. Let $f = f_k + f_{k+1} + \cdots$ be a polynomial belonging to $\mathfrak{A}$ such that $f_k \neq 0$, and let $f_k = y^\rho \psi_{k-\rho}$, $\psi_{k-\rho} \not\equiv 0(y)$. Since y^ρ and $\psi_{k-\rho}$ are relatively prime, every form g, of degree $m \geqq k-1$, can be expressed as a linear combination $A\psi_{k-\rho} + By^\rho$, where A and B are forms of degree $m-k$ and $m-\rho$ respectively. It follows that given any integer $n \geqq 0$, it is possible to find two polynomials $P^{(n)}(x,y)$, $Q^{(n)}(x,y)$ of the form

$$
\begin{aligned}
P^{(n)}(x,y) &= y^\rho + A_{\rho+1}(x,y) + \cdots + A_{\rho+n}(x,y), \\
Q^{(n)}(x,y) &= \psi_{k-\rho} + B_{k-\rho+1}(x,y) + \cdots + B_{k-\rho+n}(x,y),
\end{aligned}
\tag{3}
$$

where A_i, B_i are forms of degree i, in such a manner as to have

$$
f \equiv P^{(n)}Q^{(n)}(\mathfrak{p}^{k+n+1}). \tag{4}
$$

In fact, we have for the unknown form A_i, B_i the equations

$$
A_{\rho+i}\psi_{k-\rho} + B_{k-\rho+i}y^\rho = f_{k+i} - \sum_{j=1}^{i-1} A_{\rho+j}B_{k-\rho+i-j}, \qquad (i = 1, 2, \cdots, n),
$$

and these equations can be solved successively for $A_{\rho+1}$, $B_{k-\rho+1}$; $A_{\rho+2}$, $B_{k-\rho+2}$; etc. We take n sufficiently high, so as to have $\mathfrak{p}^{k+n+1} \equiv 0(\mathfrak{A})$.[5] For such a value of n we will have

$$
P^{(n)}Q^{(n)} \equiv 0(\mathfrak{A}). \tag{5}
$$

---

[5] Since $\mathfrak{A} = [\mathfrak{q}_i, \mathfrak{p}^k]$ and $\mathfrak{q}_i$ is a primary ideal belonging to $\mathfrak{p}$, also $\mathfrak{A}$ is a primary ideal belonging to $\mathfrak{p}$.

This implies $P^{(n)}Q^{(n)} \equiv 0(\mathfrak{q}_i)$. Now the subform $\psi_{k-\rho}$ of $Q^{(n)}$ is not divisible by y and, by our choice of the variables x, y, we have $v(y) > v(x)$. Hence $v(Q^{(n)}) = v(x^{k-\rho})$, and consequently, if $\phi_{k-\rho}$ is *any* form of degree $k-\rho$ in x, y, we have $v(\phi_{k-\rho}) \geqq v(Q^n)$. Consequently, $v(P^{(n)}\phi_{k-\rho}) \geqq v(P^{(n)}Q^{(n)})$, and, since $P^{(n)}Q^{(n)} \equiv 0(\mathfrak{q}_i)$, also $P^{(n)}\phi_{k-\rho} \equiv 0(\mathfrak{q}_i)$. This congruence holds true for any form $\phi_{k-\rho}$ of degree $k-\rho$, whence $P^{(n)}\mathfrak{p}^{k-\rho} \equiv 0(\mathfrak{q}_i)$. Since the subform $y^\rho\psi_{k-\rho}$ of $P^n\phi_{k-\rho}$ is of degree k, we have $P^{(n)}\phi_{k-\rho} \equiv 0(\mathfrak{p}^k)$, consequently, since $\mathfrak{A} = [\mathfrak{q}_i, \mathfrak{p}^k]$,

$$P^{(n)}\phi_{k-\rho} \equiv 0(\mathfrak{A}) \tag{6}$$

and

$$P^{(n)}\mathfrak{p}^{k-\rho} \equiv 0(\mathfrak{A}). \tag{6'}$$

We have then, in view of (6), that if $\mathfrak{A}$ contains a polynomial f whose subform f_k is divisible by y^ρ but not by $y^{\rho+1}$, $\mathfrak{A}$ also contains polynomials whose subforms are of degree k and are arbitrarily assigned forms divisible by y^ρ. This shows that $\Omega(\mathfrak{A})$ consists of all the forms which are divisible by a certain power of y, say y^ρ, where necessarily $\rho = k - r$, if r is the dimension of $\Omega(\mathfrak{A})$. This proves the first part of the theorem.

Let $\mathfrak{A}' = \mathfrak{A} : \mathfrak{p}^r$ and let $\mathfrak{A}' \sim \mathfrak{q}_j$, i. e. let $\mathfrak{q}_j$ be the v-ideal such that $v(\mathfrak{A}') = v(\mathfrak{q}_j)$. There exists such an ideal $\mathfrak{q}_j$ *in the sequence* $\{\mathfrak{q}_n\}$, since $\mathfrak{A}$, a primary 0-dimensional ideal, cannot belong to all the ideals $\mathfrak{q}_n$ (whose intersection $\mathfrak{p}_1$ is at least one-dimensional) and since $\mathfrak{A} \equiv 0(\mathfrak{A}')$. We have $\mathfrak{A}'\mathfrak{p}^r \equiv 0(\mathfrak{A}) \equiv 0(\mathfrak{q}_i)$, whence $\mathfrak{q}_j\mathfrak{p}^r \equiv 0(\mathfrak{q}_i)$, since $v(\mathfrak{q}_j\mathfrak{p}^r) = v(\mathfrak{A}'\mathfrak{p}^r) \geqq v(\mathfrak{q}_i)$. *We assert that* $\mathfrak{q}_j\mathfrak{p}^r$ *is also contained in the ideal* $\mathfrak{p}^k$, i. e. $\mathfrak{q}_j$ is contained in $\mathfrak{p}^{k-r}$. In fact, assume the contrary. There will then exist in $\mathfrak{q}_j$ a polynomial $F = F_\sigma + F_{\sigma+1} + \cdots$, whose subform F_σ is of degree $\sigma < k - r$, i. e. $\sigma < \rho$. The polynomial $x^{k-\sigma}F$ belongs to $\mathfrak{q}_i$, since $\mathfrak{q}_j\mathfrak{p}^r \equiv 0(\mathfrak{q}_i)$ and $k - \sigma > r$. It also belongs to $\mathfrak{p}^k$. Hence $x^{k-\sigma}F \equiv 0(\mathfrak{A})$, and this is impossible, since the subform $x^{k-\sigma}F_\sigma$ of $x^{k-\sigma}F$ is of degree k and is at most divisible by y^σ, $\sigma < \rho$. As a result, we have $\mathfrak{q}_j\mathfrak{p}^r \equiv 0(\mathfrak{q}_i)$ *and* $\mathfrak{q}_j\mathfrak{p}^r \equiv 0(\mathfrak{p}^k)$, whence

$$\mathfrak{q}_j\mathfrak{p}^r \equiv 0(\mathfrak{A}). \tag{7}$$

On the other hand, let f be any polynomial belonging to $\mathfrak{A}$, and let us first assume that its subform f_k is not divisible by $y^{\rho+1}$, $f_k = y^\rho\psi_{k-\rho}$, $\psi_{k-\rho} \not\equiv 0(y)$. As above, we determine the polynomials $P^{(n)}$ and $Q^{(n)}$, given by (3), so as to satisfy (4), and we again choose n sufficiently high so that $\mathfrak{p}^{k+n+1} \equiv 0(\mathfrak{A})$. Then the congruence (5) holds true and consequently also (6′), whence $P^{(n)} \equiv 0(\mathfrak{A}') \equiv 0(\mathfrak{q}_j)$, since $k - \rho = r$. Moreover, if n is sufficiently high, we will also have $\mathfrak{p}^{n+k+1} \equiv 0(\mathfrak{q}_j\mathfrak{p}^r)$. For such a value of n we deduce immediately from (4) that $f \equiv 0(\mathfrak{q}_j\mathfrak{p}^r)$, since we have just seen that $P^{(n)}$ belongs to $\mathfrak{q}_j$ and since $Q^{(n)} \equiv 0(\mathfrak{p}^r)$. Thus, we have shown that every polynomial f in $\mathfrak{A}$ belongs to $\mathfrak{q}_j\mathfrak{p}^r$, provided the subform of f is not divisible by $y^{\rho+1}$. If the subform f_k of f is divisible by $y^{\rho+1}$, we consider in $\mathfrak{A}$ a polynomial $\bar{f} = \bar{f}_k + \cdots$, such that $\bar{f}_k \not\equiv 0(y^{\rho+1})$. By the preceding result, we have $\bar{f} \equiv 0(\mathfrak{q}_j\mathfrak{p}^r)$ and also $\bar{f} + f \equiv 0(\mathfrak{q}_j\mathfrak{p}^r)$, whence again $f \equiv 0(\mathfrak{q}_j\mathfrak{p}^r)$. It is therefore proved that

$$\mathfrak{A} \equiv 0(\mathfrak{q}_j\mathfrak{p}^r),$$

and comparing with (7), we deduce

$$\mathfrak{A} = \mathfrak{q}_j\mathfrak{p}^r,$$

and this proves our theorem.

The following consequences can be drawn from Theorem 3:

COROLLARY 3.1. *$\mathfrak{A}$ cannot admit a factor $\mathfrak{p}^\sigma$ with $\sigma > r$ $(r = k - \rho)$,* i.e. *if $\mathfrak{A} = \mathfrak{p}^\sigma\mathfrak{A}_1$ is a product representation of $\mathfrak{A}$, then $\sigma \leqq r$.* In fact, the subforms of $\mathfrak{A}$ form then a system $\Omega(\mathfrak{A})$ of dimension $\geqq \sigma$.

COROLLARY 3.2 (special case $k = h$). *If $\mathfrak{q}_i \equiv 0(\mathfrak{p}^h)$, $\mathfrak{q}_i \not\equiv 0(\mathfrak{p}^{h+1})$ and if $\Omega(\mathfrak{q}_i)$ is of dimension r, then $\mathfrak{q}_i = \mathfrak{p}^r\mathfrak{q}_j$, where $\mathfrak{q}_j$ is an ideal in the sequence $\{\mathfrak{q}_n\}$, and $\mathfrak{q}_i$ does not admit as a factor a higher power of $\mathfrak{p}$ than $\mathfrak{p}^r$. In particular, if $r = 0$, then $\mathfrak{q}_i$ does not admit factors $\mathfrak{p}$ and every element of $\Omega(\mathfrak{q}_i)$ coincides, to within a constant factor in $\mathfrak{k}$, with y^h.*

COROLLARY 3.3. *If $\mathfrak{q}_i \equiv 0(\mathfrak{p}^h)$, $\mathfrak{q}_i \not\equiv 0(\mathfrak{p}^{h+1})$ and if $\mathfrak{p}^k\mathfrak{q}_i \sim \mathfrak{q}_m$ $(k \geqq 0)$, then*

$$[\mathfrak{q}_m, \mathfrak{p}^{h+k}] = \mathfrak{p}^k\mathfrak{q}_i.$$

In fact, let $\mathfrak{A} = [\mathfrak{q}_m, \mathfrak{p}^{h+k}]$. We have $\mathfrak{q}_m \not\equiv 0(\mathfrak{p}^{h+k+1})$, hence, by Theorem 3, $\mathfrak{A} = \mathfrak{p}^r\mathfrak{q}_j$, where $\mathfrak{q}_j$ is some v-ideal of our sequence $\{\mathfrak{q}_n\}$ and r is the dimension of $\Omega(\mathfrak{A})$. Since $\mathfrak{p}^k\mathfrak{q}_i \sim \mathfrak{q}_m$, we have $\mathfrak{p}^k\mathfrak{q}_i \equiv 0(\mathfrak{q}_m)$ and also $\mathfrak{p}^k\mathfrak{q}_i \equiv 0(\mathfrak{p}^{h+k})$, since $\mathfrak{q}_i \equiv 0(\mathfrak{p}^h)$. Consequently $\mathfrak{p}^k\mathfrak{q}_i \equiv 0(\mathfrak{A}) \equiv 0(\mathfrak{p}^r\mathfrak{q}_j)$. Since the dimension of $\Omega(\mathfrak{p}^k\mathfrak{q}_i)$ is at least k and since $\Omega(\mathfrak{p}^k\mathfrak{q}_i)$ is *a* subset of $\Omega(\mathfrak{A})$ (in view of the assumption $\mathfrak{q}_i \not\equiv 0(\mathfrak{p}^{h+1})$), it follows $k \leqq r$. Now, $\mathfrak{q}_m \sim \mathfrak{p}^k\mathfrak{q}_i$,

$$v(\mathfrak{p}^k\mathfrak{q}_i) = v(\mathfrak{q}_m) \leqq v(\mathfrak{A}) = v(\mathfrak{p}^r\mathfrak{q}_j), \text{ i.e. } v(\mathfrak{q}_i) \leqq v(\mathfrak{p}^{r-k}\mathfrak{q}_j),$$

whence $\mathfrak{p}^{r-k}\mathfrak{q}_j \equiv 0(\mathfrak{q}_i)$ and $\mathfrak{p}^r\mathfrak{q}_j \equiv 0(\mathfrak{p}^k\mathfrak{q}_i)$. Since we also have $\mathfrak{p}^k\mathfrak{q}_i \equiv 0(\mathfrak{p}^r\mathfrak{q}_j)$, it follows that $\mathfrak{A} = \mathfrak{p}^r\mathfrak{q}_j = \mathfrak{p}^k\mathfrak{q}_i$, *q. e. d.*

4. v-ideals and quadratic transformations. We consider the quadratic transformation T:

$$x' = x,\ y' = y/x; \qquad x = x',\ y = x'y',$$

having at $x = y = 0$ a fundamental point, and we denote by $\mathfrak{O}'$ the ring of polynomials in x', y'. $\mathfrak{O}$ is a subring of $\mathfrak{O}'$, and moreover $\mathfrak{O}'$ is contained in the valuation ring of the given valuation B, since, we have assumed $v(y) > v(x)$, whence $v(y') > 0$. Let $\{\mathfrak{q}'_j\}$ be the sequence of 0-dimensional v-ideals in $\mathfrak{O}'$, where $\mathfrak{q}'_1 = \mathfrak{p}' = (x', y')$. We wish to study the connection between the v-ideals $\mathfrak{q}_i$ in $\mathfrak{O}$ and the v-ideals $\mathfrak{q}'_j$ in $\mathfrak{O}'$. Note that $\mathfrak{q}_1 = \mathfrak{p} = (x, y)$, $\mathfrak{q}_2 = (y, x^2)$, whence $\mathfrak{O}'\mathfrak{p} = (x', x'y') = (x')$, where $(x') = \mathfrak{O}'x'$, and $\mathfrak{O}'\mathfrak{q}_2 = (x'y', x'^2) = x'\mathfrak{p}'$.

THEOREM 4.1. *The extended ideal of an ideal $\mathfrak{q}$ in the sequence $\{\mathfrak{q}_i\}$ is of the form $x'^h\mathfrak{q}'$, where $\mathfrak{q}'$ is an ideal in the sequence $\{\mathfrak{q}'_j\}$ and where $\mathfrak{q} \equiv 0(\mathfrak{p}^h)$, $\mathfrak{q} \not\equiv 0(\mathfrak{p}^{h+1})$. Moreover, $\mathfrak{q}$ is the contracted ideal of $x'^h\mathfrak{q}'$.*

Proof. If $f(x, y) = f_\sigma(x, y) + f_{\sigma+1}(x, y) + \cdots$ is a polynomial in x, y and $f_\sigma(x, y)$ is its subform, then

$$f(x, y) = f(x', x'y) = x'^\sigma f_\sigma(1, y') + x'^{\sigma+1} f_{\sigma+1}(1, y') + \cdots,$$

whence, considering f as an element of $\mathfrak{O}'$, we have $f \equiv 0(x'^\sigma)$, $f \not\equiv 0(x'^{\sigma+1})$. By hypothesis, the subform of any polynomial belonging to $\mathfrak{q}$ is of degree $\geqq h$ and $\mathfrak{q}$ contains a polynomial whose subform is exactly of degree h. Hence $\mathfrak{O}'\mathfrak{q} = x'^h\mathfrak{q}'$, where $\mathfrak{q}' \not\equiv 0(x')$.

We show that $\mathfrak{q}'$ is a v-ideal. We have

$$v(\mathfrak{q}') = v(\mathfrak{q}) - hv(x). \tag{8}$$

Let ω be an element of $\mathfrak{O}'$ such that $v(\omega) \geqq v(\mathfrak{q}')$, and let

$$\omega = F(x', y') = F(x, y/x) = \frac{G(x, y)}{x^\sigma},$$

where F and G are polynomials and $G(x, y) \equiv 0(\mathfrak{p}^\sigma)$ (since clearly the subform of G is of degree $\geqq \sigma$). Since $v(\omega) \geqq v(\mathfrak{q}')$ it follows, by (8), $(h - \sigma)v(x) + v(G) \geqq v(\mathfrak{q})$. Let k be a non-negative integer such that

$$h_1 = h - \sigma + k \geqq 0.$$

We will have then

$$v(x^{h_1}G) \geqq v(x^k\mathfrak{q}) = v(\mathfrak{p}^k\mathfrak{q}). \tag{9}$$

Let $\mathfrak{p}^k\mathfrak{q} \sim \mathfrak{q}_m$, where $\mathfrak{q}_m$ belongs to the sequence $\{\mathfrak{q}_i\}$ of v-ideals in $\mathfrak{O}$. The inequality (9) implies $x^{h_1}G \equiv 0(\mathfrak{q}_m)$, and since $x^{h_1}G \equiv 0(\mathfrak{p}^{h_1+\sigma}) \equiv 0(\mathfrak{p}^{h+k})$, we have

$$x^{h_1}G \equiv 0([\mathfrak{q}_m, \mathfrak{p}^{h+k}]).$$

By Corollary 3.3 of the preceding section it follows that

$$x^{h_1}G \equiv 0(\mathfrak{q}\mathfrak{p}^k),$$

whence $x^{h_1}G$ is of the form

$$x^{h_1}G = \Sigma A_i(x, y) B_i(x, y),$$

where $A_i \equiv 0(\mathfrak{q})$, $B_i \equiv 0(\mathfrak{p}^k)$. But then B_i/x^k is a polynomial in x', y', and putting $B'_i = B_i/x^k$ we have

$$x^{h-\sigma}G = x^{h_1}G/x^k = \Sigma A_i B'_i \equiv 0(\mathfrak{O}'\mathfrak{q}) \equiv 0(x'^h\mathfrak{q}'),$$

whence $\omega = G/x^\sigma \equiv 0(\mathfrak{q}')$. We have thus proved that the ideal $\mathfrak{q}'$ enjoys the property which characterizes the v-ideals: if $v(\omega) \geqq v(\mathfrak{q}')$, then $\omega \equiv 0(\mathfrak{q}')$.

It remains to prove that $\mathfrak{q}'$ belongs to the sequence $\{\mathfrak{q}'_j\}$. But this is obvious, since $\mathfrak{q}'$ is necessarily a 0-dimensional ideal[6] or the unit ideal. That the contracted ideal $\bar{\mathfrak{q}}$ of $x'^h\mathfrak{q}'$ coincides with $\mathfrak{q}$ follows immediately from the fact that $v(\bar{\mathfrak{q}}) \geqq v(x'^h\mathfrak{q}') = v(\mathfrak{q})$, whence $\bar{\mathfrak{q}} \equiv 0(\mathfrak{q})$, while on the other hand we must have, of course, $\mathfrak{q} \equiv 0(\bar{\mathfrak{q}})$. The theorem is thus proved.

The next theorem is in a sense the converse of the preceding theorem.

THEOREM 4.2. *For any ideal $\mathfrak{q}'$ in the sequence $\{\mathfrak{q}'_j\}$ there exists an integer h such that $x'^h\mathfrak{q}'$ is the extended ideal of an ideal $\mathfrak{q}$ belonging to the sequence $\{\mathfrak{q}_i\}$.*

Proof. If $\phi_1(x', y')$, $\phi_2(x', y'), \cdots, \phi_k(x', y')$ is a base of $\mathfrak{q}'$ and if we write these polynomials in the form of quotients: $\phi_i(x', y') = \psi_i(x, y)/x^m$, with a common denominator x^m, where $\psi_1, \psi_2, \cdots, \psi_k$ are polynomials in x, y, then we see that $x'^m\mathfrak{q}'$ is the extended ideal of the ideal $(\psi_1, \psi_2, \cdots, \psi_k)$ in $\mathfrak{O}$. Thus, there exist integers m such that $x'^m\mathfrak{q}'$ is an extended ideal of an ideal in $\mathfrak{O}$. This will be true for all sufficiently high integers m, since if $\mathfrak{O}'\bar{\mathfrak{q}} = x'^m\mathfrak{q}'$, then $\mathfrak{O}'\mathfrak{p}\bar{\mathfrak{q}} = x'^{m+1}\mathfrak{q}'$. Let h be the smallest possible value of m, and let $\mathfrak{q}$ be the contracted ideal of $x'^h\mathfrak{q}$:

$$\mathfrak{q} = [\mathfrak{O}, x'^h\mathfrak{q}'], \qquad \mathfrak{O}'\mathfrak{q} = x'^h\mathfrak{q}'. \tag{10}$$

Since the contracted ideal of (x'^h) is $\mathfrak{p}^h$, it follows $\mathfrak{q} \equiv 0(\mathfrak{p}^h)$. Moreover, since the primary ideal $\mathfrak{q}'$ belongs to the prime ideal $\mathfrak{p}' = (x', y')$, we have $x'^n \equiv 0(x'^h\mathfrak{q}')$, if n is sufficiently high. Passing to the contracted ideals we find $\mathfrak{p}^n \equiv 0(\mathfrak{q})$. Hence there exists an ideal $\mathfrak{q}_n$ in the sequence $\{\mathfrak{q}_i\}$ of v-ideals in $\mathfrak{O}$ such that $v(\mathfrak{q}_n) = v(\mathfrak{q})$. Let $\mathfrak{p}^\sigma$ be the highest power of $\mathfrak{p}$ which divides

---

[6] If n is a sufficiently high integer, then $x^{n+h} \equiv 0(\mathfrak{q})$, whence $x'^n \equiv 0(\mathfrak{q}')$. Let $f(x, y)$ be a polynomial in $\mathfrak{q}$ whose subform is $f_h(x, y)$, $f_h \neq 0$. If $f = f_h + f_{h+1} + \cdots$, then the polynomial $f_h(1, y') + x' f_{h+1}(1, y') + \cdots$ belongs to $\mathfrak{q}'$. It follows that also $[f_h(1, y')]^n$ belongs to $\mathfrak{q}'$, hence $(x'^n, [f_h(1, y')]^n) \equiv 0(\mathfrak{q}')$, and consequently $\mathfrak{q}'$ is 0-dimensional, or is the unit ideal, since $f_h(1, y') \neq 0$.

$\mathfrak{q}_n$, $\mathfrak{q}_n \equiv 0(\mathfrak{p}^\sigma)$, $\mathfrak{q}_n \not\equiv 0(\mathfrak{p}^{\sigma+1})$. We have $\mathfrak{q} \equiv 0(\mathfrak{p}^h)$ and $\mathfrak{q} \not\equiv 0(\mathfrak{p}^{h+1})$, consequently $\sigma \leqq h$, because $\mathfrak{q} \equiv 0(\mathfrak{q}_n)$. Let $\mathfrak{O}'\mathfrak{q}_n = x'^\sigma \mathfrak{q}'_j$, where $\mathfrak{q}'_j$, by the preceding theorem, is a v-ideal in $\mathfrak{O}'$. Since $\mathfrak{q}_n \sim \mathfrak{q}$, we have $v(x'^\sigma \mathfrak{q}'_j) = v(x'^h \mathfrak{q}')$. If $\sigma = h$, then $v(\mathfrak{q}'_j) = v(\mathfrak{q}')$, whence $\mathfrak{q}'_j = \mathfrak{q}'$, since both are v-ideals. In this case $\mathfrak{q}_n$ and $\mathfrak{q}'$ must coincide, since both are contracted ideals of $x'^h \mathfrak{q}'$, and the theorem is proved.

Assume $\sigma < h$. We have

$$\mathfrak{q} \equiv 0([\mathfrak{q}_n, \mathfrak{p}^h]). \tag{11}$$

If $f(x, y)$ is any polynomial belonging to $[\mathfrak{q}_n, \mathfrak{p}^h]$, then $f = x'^h f'(x'y')$ and $f \equiv 0(x'^\sigma \mathfrak{q}'_j)$. Since $v(x'^\sigma \mathfrak{q}'_j) = v(x'^h \mathfrak{q}')$, it follows that $f'(x', y') \equiv 0(\mathfrak{q}')$. Hence $f \equiv 0(x'^h \mathfrak{q}')$ and consequently, by (10), $f \equiv 0(\mathfrak{q})$. We have therefore $[\mathfrak{q}_n, \mathfrak{p}^h] \equiv 0(\mathfrak{q})$, and, by (11), it follows

$$\mathfrak{q} = [\mathfrak{q}_n, \mathfrak{p}^h]. \tag{12}$$

We have $\mathfrak{q}_n \equiv 0(\mathfrak{p}^\sigma)$, $\mathfrak{q}_n \not\equiv 0(\mathfrak{p}^{\sigma+1})$ and $h > \sigma$. We can then apply Theorem 3 and we obtain $\mathfrak{q} = \mathfrak{p}^r \mathfrak{q}_a$, where $\mathfrak{q}_a$ is again a v-ideal in $\mathfrak{O}$. Here r is the dimension of $\Omega(\mathfrak{q})$ and consequently $r > 0$, because $\mathfrak{p}^{h-\sigma}\mathfrak{q}_n \equiv 0(\mathfrak{q})$, whence $r \geqq h - \sigma > 0$. We have then

$$\mathfrak{O}'\mathfrak{q} = \mathfrak{O}'\mathfrak{p}^r \mathfrak{q}_a = x'^r \mathfrak{O}'\mathfrak{q}_a = x'^h \mathfrak{q}',$$

whence $\mathfrak{O}'\mathfrak{q}_a = x'^{h-r}\mathfrak{q}'$. This contradicts our hypothesis that h is the smallest integer such that $x'^h \mathfrak{q}'$ is an extended ideal. Hence $\sigma < h$ is impossible, and the theorem is proved.

If $\mathfrak{A}$ is any ideal in $\mathfrak{O}$ and if $\mathfrak{O}'\mathfrak{A} = x'^h \mathfrak{A}'$, $\mathfrak{A}' \not\equiv 0(x')$, we shall call $\mathfrak{A}'$ the *transformed ideal* of $\mathfrak{A}$ (under the quadratic transformation T), in symbols: $\mathfrak{A}' = T(\mathfrak{A})$. By Theorems 4.1 and 4.2, the transform of any v-ideal $\mathfrak{q}$ in $\mathfrak{O}$ is a v-ideal $\mathfrak{q}'$ in $\mathfrak{O}'$, and every v-ideal $\mathfrak{q}'$ in $\mathfrak{O}'$ is the transform of at least one v-ideal $\mathfrak{q}$ in $\mathfrak{O}$, everything referred to a fixed valuation. There may be more than one v-ideal in $\mathfrak{O}$ whose transform is $\mathfrak{q}'$. To find them, we again consider the smallest integer h such that $x'^h \mathfrak{q}'$ is an extended ideal. It follows from the preceding proof that if $\mathfrak{q}$ is the contracted ideal of $x'^h \mathfrak{q}'$, then $T(\mathfrak{q}) = \mathfrak{q}'$, and from Theorem 4.1 it follows, that any other v-ideal $\bar{\mathfrak{q}}$ in $\mathfrak{O}$ such that $T(\bar{\mathfrak{q}}) = \mathfrak{q}'$ must be the contracted ideal of $x'^\sigma \mathfrak{q}'$, where σ is some integer greater than h. For a given integer σ greater than h, the contracted ideal of $x'^\sigma \mathfrak{q}'$ may or may not be a v-ideal in $\mathfrak{O}$, but let us at any rate examine this contracted ideal. Let us denote it by $\mathfrak{A}_\sigma$. Since $x'^\sigma \mathfrak{q}'$ is the extended ideal of $\mathfrak{A}_\sigma$ and *also* of $\mathfrak{p}^{\sigma-h}\mathfrak{q}$, it follows

$$\mathfrak{p}^{\sigma-h}\mathfrak{q} \equiv 0(\mathfrak{A}_\sigma). \tag{13}$$

For the same reason we have $v(\mathfrak{A}_\sigma) = v(\mathfrak{p}^{\sigma-h}\mathfrak{q})$. Let $\mathfrak{q}_m$ be the v-ideal equivalent to both $\mathfrak{A}_\sigma$ and $\mathfrak{p}^{\sigma-h}\mathfrak{q}$, $\mathfrak{q}_m \sim \mathfrak{A} \sim \mathfrak{p}^{\sigma-h}\mathfrak{q}$. By Theorem 3, Corollary 3. 3, we have $\mathfrak{p}^{\sigma-h}\mathfrak{q} = [\mathfrak{q}_m, \mathfrak{p}^\sigma]$. Now, $\mathfrak{A}_\sigma \sim \mathfrak{q}_m$ implies $\mathfrak{A}_\sigma \equiv 0(\mathfrak{q}_m)$ and since $\mathfrak{O}'\mathfrak{A}_\sigma = x'^\sigma\mathfrak{q}'$ we also have $\mathfrak{A}_\sigma \equiv 0(\mathfrak{p}^\sigma)$. Consequently $\mathfrak{A}_\sigma \equiv 0(\mathfrak{p}^{\sigma-h}\mathfrak{q})$, and hence, by (13), $\mathfrak{A}_\sigma = \mathfrak{p}^{\sigma-h}\mathfrak{q}$. We therefore can state the following theorem.

THEOREM 4. 3. *If $\mathfrak{q}'$ is a v-ideal in $\mathfrak{O}'$, belonging to the sequence $\{\mathfrak{q}'_j\}$, and if h is the smallest integer such that $x'^h\mathfrak{q}'$ is an extended ideal of an ideal in $\mathfrak{O}$, then the contracted ideal $\mathfrak{q}$ of $x'^h\mathfrak{q}'$ is a v-ideal in $\mathfrak{O}$, a member of the sequence $\{\mathfrak{q}_i\}$, and $T(\mathfrak{q}) = \mathfrak{q}'$. If σ is any integer $\geqq h$, the contracted ideal of $x'^\sigma\mathfrak{q}'$ is $\mathfrak{p}^{\sigma-h}\mathfrak{q}$, but need not be a v-ideal. In particular, the v-ideals $\mathfrak{q}_i$ whose transform is the given v-ideal $\mathfrak{q}'$ are all of the form $\mathfrak{p}^{\sigma-h}\mathfrak{q}$, $\sigma \geqq h$.*

We shall regard $\mathfrak{q}$ as the transform of $\mathfrak{q}'$ by T^{-1}: $\mathfrak{q} = T^{-1}(\mathfrak{q}')$. By the definition of $\mathfrak{q}$, the system $\Omega(\mathfrak{q})$ of the subforms of $\mathfrak{q}$ must be of dimension $r = 0$, i. e. every form in $\Omega(\mathfrak{q})$ differs from y^h by a factor c, $c \varepsilon \mathfrak{k}$. In fact, if it were $r > 0$, then $\mathfrak{q}$ could be put in the form $\mathfrak{q} = \mathfrak{p}^r\bar{\mathfrak{q}}$, where $\bar{\mathfrak{q}}$ is also a v-ideal (Corollary 3. 2), and we would have $\mathfrak{O}'\bar{\mathfrak{q}} = x'^{h-r}\mathfrak{q}'$, contrary to our assumption that h is the smallest integer such that $x'^h\mathfrak{q}'$ is an extended ideal. Conversely, if $\mathfrak{q}$ is a v-ideal in $\mathfrak{O}$, belonging to the sequence $\{\mathfrak{q}_i\}$, and if $\Omega(\mathfrak{q})$ is of dimension zero, then $\mathfrak{q}$ is so related to its transform $\mathfrak{q}' = T(\mathfrak{q})$, that $\mathfrak{q} = T^{-1}\mathfrak{q}'$, i. e. if $\mathfrak{O}'\mathfrak{q} = x'^h\mathfrak{q}'$, then h is the smallest integer such that $x'^h\mathfrak{q}'$ is an extended ideal. In fact, in the contrary case, $\mathfrak{q}$ could be put in the form $\mathfrak{p}^r\bar{\mathfrak{q}}$, $r > 0$, (by Theorem 4. 3), contrary to the hypothesis that $\Omega(\mathfrak{q})$ is of dimension zero. Thus, there is a one to one correspondence between the ideals $\mathfrak{q}'$ of the sequence $\{\mathfrak{q}'_j\}$ and those ideals $\mathfrak{q}$ of the sequence $\{\mathfrak{q}_i\}$ whose system $\Omega(\mathfrak{q})$ of subforms is of dimension zero: to each $\mathfrak{q}'$ there corresponds a unique ideal $\mathfrak{q} = T^{-1}(\mathfrak{q}')$, and $\mathfrak{q}' = T(\mathfrak{q})$.

Let $\mathfrak{q}'_\alpha$, $\mathfrak{q}'_\beta$ be two distinct v-ideals in $\mathfrak{O}'$, and let $\mathfrak{q}_i = T^{-1}(\mathfrak{q}'_\alpha)$, $\mathfrak{q}_j = T^{-1}(\mathfrak{q}'_\beta)$, be their transforms in $\mathfrak{O}$. Suppose that $i < j$, whence $\mathfrak{q}_j \equiv 0(\mathfrak{q}_i)$. *We assert that in such a case also $\alpha < \beta$.* In fact, let $\mathfrak{q}_i \equiv 0(\mathfrak{p}^h)$, $\mathfrak{q}_i \not\equiv 0(\mathfrak{p}^{h+1})$ and let $\mathfrak{q}_j \equiv 0(\mathfrak{p}^\sigma)$, $\mathfrak{q}_j \not\equiv 0(\mathfrak{p}^{\sigma+1})$, whence $\mathfrak{O}'\mathfrak{q}_i = x'^h\mathfrak{q}'_\alpha$, $\mathfrak{O}'\mathfrak{q}_j = x'^\sigma\mathfrak{q}'_\beta$, and clearly, $\sigma \geqq h$. Evidently $\mathfrak{p}^{\sigma-h}\mathfrak{q}_i \equiv 0(x'^\sigma\mathfrak{q}'_\alpha)$. Supposing that $\alpha > \beta$, whence $\mathfrak{q}'_\alpha \equiv 0(\mathfrak{q}'_\beta)$, we would have $x'^\sigma\mathfrak{q}'_\alpha \equiv 0(x'^\sigma\mathfrak{q}'_\beta)$, and passing to the contracted ideals in $\mathfrak{O}$, we would get by Theorem 4. 3 $\mathfrak{p}^{\sigma-h}\mathfrak{q}_i \equiv 0(\mathfrak{q}_j)$. The equality $\sigma = h$ is excluded, because $\mathfrak{q}_j \equiv 0(\mathfrak{q}_i)$ and $\mathfrak{q}_j \neq \mathfrak{q}_i$. Hence $\sigma > h$, but then the congruence $\mathfrak{p}^{\sigma-h}\mathfrak{q}_i \equiv 0(\mathfrak{q}_j)$ is in contradiction with the fact that the $\Omega(\mathfrak{q}_j)$ (consisting of form of degree σ) is of dimension zero. Hence our assumption $\alpha > \beta$ leads to a contradiction, and consequently it is proved that $i < j$ implies $\alpha < \beta$. If then $T^{-1}(\mathfrak{q}'_j) = \mathfrak{q}_{\alpha_j}$, the indices α_j form an ascending sequence, $\alpha_0 < \alpha_1 < \alpha_2 < \cdots$. It is immediately verified that $\alpha_0 = 0$, $\alpha_1 = 2$. Hence $\alpha_j > j$, if $j > 0$. Let now $\mathfrak{q}_s$ be any v-ideal in the sequence $\{\mathfrak{q}_i\}$ and let $T(\mathfrak{q}_s) = \mathfrak{q}'_\sigma$. By Theorem 4. 3 we have $\mathfrak{q}_s = \mathfrak{p}^\rho \mathfrak{q}_{\alpha_\sigma}, \rho \geqq 0$, whence $\mathfrak{q}_s \equiv 0(q_{\alpha_\sigma})$, and $s \geqq \alpha_\sigma$, i. e. $s > \sigma$. Observing that the length of the ideal $\mathfrak{q}_s$ is equal to s, we can reassume the preceding results in the following theorem:

THEOREM 4. 4. *If* $\mathfrak{q}_{\alpha_j} = T^{-1}(\mathfrak{q}'_j)$, *then* $\alpha_j > j$, *if* $j \geqq 1$. *Moreover, if* $\mathfrak{q}_s$ *is any ideal in the sequence* $\{\mathfrak{q}_i\}$ *and if* $\mathfrak{q}'_\sigma = T(\mathfrak{q}_s)$, *then* $s > \sigma$, *i. e. length of* $\mathfrak{q}_s >$ *length of* $\mathfrak{q}_\sigma$.

5. Simple and composite *v*-ideals. We say that an ideal $\mathfrak{A}$ in $\mathfrak{O}$ is *simple,* if $\mathfrak{A}$ cannot be represented as the product of two ideals, both different from the unit ideal, i. e. if $\mathfrak{A} = \mathfrak{B}\mathfrak{C}$, $\mathfrak{B} \neq (1)$ implies $\mathfrak{C} = (1)$. An ideal is *composite* if it is not simple.

THEOREM 5. 1. *A composite v-ideal can be represented as a product of v-ideals different from the unit ideal.*

Proof. Let $\mathfrak{A}$ be a v-ideal in $\mathfrak{O}$, belonging to some valuation B, and let $\mathfrak{A} = \mathfrak{B}\mathfrak{C}$, $\mathfrak{B} \neq (1)$, $\mathfrak{C} \neq (1)$. Let $\mathfrak{B}_1$, $\mathfrak{C}_1$ be the v-ideals belonging to B such that $\mathfrak{B} \sim \mathfrak{B}_1$, $\mathfrak{C} \sim \mathfrak{C}_1$. Since $\mathfrak{B} \equiv 0(\mathfrak{B}_1)$, $\mathfrak{C} \equiv 0(\mathfrak{C}_1)$, we have $\mathfrak{A} \equiv 0(\mathfrak{B}_1\mathfrak{C}_1)$. On the other hand, $v(\mathfrak{A}) = v(\mathfrak{B}) + v(\mathfrak{C}) = v(\mathfrak{B}_1) + v(\mathfrak{C}_1) = v(\mathfrak{B}_1\mathfrak{C}_1)$, whence $\mathfrak{B}_1\mathfrak{C}_1 \equiv 0(\mathfrak{A})$, since $\mathfrak{A}$ is a v-ideal. We conclude that $\mathfrak{A} = \mathfrak{B}_1\mathfrak{C}_1$, and it remains to prove that $\mathfrak{B}_1 \neq (1)$ and $\mathfrak{C}_1 \neq (1)$. Assume the contrary, and let, for instance $\mathfrak{B}_1 = (1)$. Then $\mathfrak{A} = \mathfrak{C}_1$, whence $\mathfrak{C} \equiv 0(\mathfrak{A})$, and consequently $\mathfrak{C} = \mathfrak{A}$ since $\mathfrak{A} = \mathfrak{B}\mathfrak{C} \equiv 0(\mathfrak{C})$. We have then $\mathfrak{A} = \mathfrak{B}\mathfrak{A}$, and this implies ([2], p. 36), $\mathfrak{B} = (1)$, contrary to hypothesis.

Consider the given valuation B and the corresponding sequences $\{\mathfrak{q}_i\}$, $\{\mathfrak{q}'_j\}$ of 0-dimensional v-ideals in the polynomial rings $\mathfrak{O} = \mathfrak{k}[x,y]$, $\mathfrak{O}' = \mathfrak{k}[x',y']$ respectively. Let, as before, $T^{-1}(\mathfrak{q}'_j) = \mathfrak{q}_{\alpha_j}$. It is clear that all the simple v-ideals of the sequence $\{\mathfrak{q}_i\}$, except the ideal $\mathfrak{p} = \mathfrak{q}_1$, belong to the sequence $\{\mathfrak{q}_{\alpha_j}\}$, since any ideal $\mathfrak{q}_i$ $(i \neq 1)$, not in the sequence $\{\mathfrak{q}_{\alpha_j}\}$, is, by Theorem 4. 3, either of the form $\mathfrak{p}^\rho \mathfrak{q}_{\alpha_j}$, $\rho > 0$, $\mathfrak{q}_{\alpha_j} \neq (1)$, or of the form $\mathfrak{p}^\rho$, $\rho > 1$. Now suppose that $\mathfrak{q}_{\alpha_j}$, for a given j, is a composite ideal. Then we can write, by Theorem 5. 1, $\mathfrak{q}_{\alpha_j} = \mathfrak{q}_s \mathfrak{q}_t$, where $\mathfrak{q}_s, \mathfrak{q}_t$ are in the sequence $\{\mathfrak{q}_i\}$. Hence $\mathfrak{q}'_j = T(\mathfrak{q}_{\alpha_j}) = \mathfrak{q}'_\sigma \mathfrak{q}'_\tau$, where $\mathfrak{q}'_\sigma = T(\mathfrak{q}_s)$ and $\mathfrak{q}'_\tau = T(\mathfrak{q}_t)$. We have seen above that $\Omega(\mathfrak{q}_{\alpha_j})$ is of dimension zero; therefore neither $\mathfrak{q}_s$ nor $\mathfrak{q}_t$ can be a power of $\mathfrak{p}$. It follows that $\mathfrak{q}'_\sigma \neq (1)$ and $\mathfrak{q}'_\tau \neq (1)$, i. e. $\mathfrak{q}'_j$ is composite. We conclude then that if $\mathfrak{q}'_j$ is a simple v-ideal in $\mathfrak{O}'$, then its transform $\mathfrak{q}_{\alpha_j} = T^{-1}(\mathfrak{q}'_j)$ is also a simple v-ideal. Much more difficult is to prove the converse:

THEOREM 5. 2. *The transform* $T(\mathfrak{q}_i)$ *of a simple v-ideal* $\mathfrak{q}_i$ *in* $\mathfrak{O}$ *is a simple v-ideal (in* $\mathfrak{O}'$*).*

The proof of this theorem will be given in Part II of the paper (Corollary 11. 2), where we shall characterize the simple v-ideals from the point of view of formal power series. Here we shall use this theorem without proof.

Let $\mathfrak{P}_1, \mathfrak{P}_2, \cdots, \mathfrak{P}_i, \cdots$ be the sequence of simple v-ideals, different from (1), as they occur in the sequence $\{\mathfrak{q}_i\}$:

$$(14) \qquad \mathfrak{P}_1 \supset \mathfrak{P}_2 \supset \cdots, \qquad \mathfrak{P}_1 = \mathfrak{p} = (x, y), \ \mathfrak{P}_2 = \mathfrak{q}_2 = (y, x^2).$$

Let similarly $\mathfrak{P}'_1, \mathfrak{P}'_2, \cdots$ be the sequence of simple v-ideals, different from (1), as they occur in the sequence $\{\mathfrak{q}'_\nu\}$. By the preceding results, especially by Theorem 5. 2, there is (1, 1) correspondence between the ideals $\mathfrak{P}_i$, $i > 1$, and the ideals $\mathfrak{P}'_\alpha$, where to $\mathfrak{P}_i$ corresponds $T(\mathfrak{P}_i) = \mathfrak{P}'_\alpha$, and $\mathfrak{P}_i = T^{-1}(\mathfrak{P}'_\alpha)$. Moreover, by Theorem 4. 4, if $T(\mathfrak{P}_i) = \mathfrak{P}'_\alpha$, and $T(\mathfrak{P}_j) = \mathfrak{P}'_\beta$, then $i < j$ implies $\alpha < \beta$. Since $T(\mathfrak{P}_2) = \mathfrak{P}'_1$, we conclude with the following theorem:

THEOREM 5. 3. $T(\mathfrak{P}_i) = \mathfrak{P}'_{i-1}$, *i. e. the transform of the simple v-ideal* $\mathfrak{P}_i$ *by the quadratic transformation T is the simple v-ideal* $\mathfrak{P}'_{i-1}$.

As a consequence of this theorem, it follows incidentally that the sequence $\{\mathfrak{q}_n\}$ contains infinitely many simple v-ideals. In fact, if we assume that any sequence $\{\mathfrak{q}_n\}$ of v-ideals in any polynomial ring contains always at least $k > 0$ simple v-ideals (it always contains at least one, namely the prime ideal $\mathfrak{q}_1 = \mathfrak{p} = (x, y)$), and if we apply this assumption to the sequence $\{\mathfrak{q}'_\nu\}$ of v-ideals in the polynomial ring $\mathfrak{O}'$, we deduce immediately, by Theorem 5. 3, that any sequence $\{\mathfrak{q}_n\}$ contains at least $k + 1$ simple v-ideals.

6. Properties of simple *v*-ideals. Heretofore we have been dealing with a fixed valuation B and with the v-ideals in $\mathfrak{O}$ belonging to B. Now, a v-ideal belonging to B may also occur as a v-ideal for many other valuations. Consider, in particular, the i-th simple v-ideal $\mathfrak{P}_i$ for B, and let $\bar{B}$ be another valuation for which $\mathfrak{P}_i$ is a v-ideal. We assert that $\mathfrak{P}_i$ *is also the i-th simple v-ideal for* $\bar{B}$. The assertion is trivial for $i = 1$, because $\mathfrak{P}_1 = \mathfrak{p} = (x, y)$. We may then proceed by induction, assuming that our assertion is true for $i - 1$. Let $\{\mathfrak{q}_n\}$ and $\{\bar{\mathfrak{q}}_n\}$ be the sequences of v-ideals in $\mathfrak{O}$ for the valuations B and $\bar{B}$ respectively. Since $\mathfrak{P}_i$ occurs in both sequences, and since the ideals $\mathfrak{q}_n$ and $\bar{\mathfrak{q}}_n$ are primary ideals belonging to the prime 0-dimensional ideals $\mathfrak{q}_1$ and $\bar{\mathfrak{q}}_1$ respectively, it follows that $\mathfrak{q}_1 = \bar{\mathfrak{q}}_1 = \mathfrak{p} = (x, y)$. Furthermore, since $\mathfrak{P}_i$ is simple, its system $\Omega(\mathfrak{P}_i)$ of subforms is of dimension 0. If $cx + dy$ is the base of $\Omega(\mathfrak{P}_i)$, we must have $v(cx + dy) > v(\mathfrak{p})$ in B and also

$v(cx + dy) > v(\mathfrak{p})$ in $\bar{B}$. We may therefore assume that $v(y) > v(x)$ in both valuations B and $\bar{B}$. We apply the quadratic transformation $T: x' = x$, $y' = y/x$ to both sequences $\{\mathfrak{q}_n\}$ and $\{\bar{\mathfrak{q}}_n\}$. The sequences $\{\mathfrak{q}'_\nu\}$ and $\{\bar{\mathfrak{q}}'_\nu\}$ of v-ideals in $\mathfrak{O}' = \mathfrak{k}[x', y']$ belonging to the valuations B and $\bar{B}$ respectively, will consist of primary ideals belonging to the prime ideal $\mathfrak{p}' = (x', y')$. The transform $\mathfrak{P}'_{i-1}$ of $\mathfrak{P}_i$ belongs to both sequences and is the $(i-1)$-th simple v-ideal in the sequence $\{\mathfrak{q}'_\nu\}$. Hence, by our induction, $\mathfrak{P}'_{i-1}$ is also the $(i-1)$-th simple ideal in the sequence $\{\bar{\mathfrak{q}}'_\nu\}$. As a consequence, $\mathfrak{P}_i$ must be the i-th simple ideal in the sequence $\{\bar{\mathfrak{q}}_n\}$, and this proves our assertion.

Thus, given a simple v-ideal $\mathfrak{P} = \mathfrak{P}_i$ in $\mathfrak{O}$, there is uniquely determined an integer i, such that $\mathfrak{P}$ is the i-th simple v-ideal in the sequence of simple v-ideals of any valuation for which $\mathfrak{P}$ is a valuation ideal. We shall say that $\mathfrak{P}$ is *a simple v-ideal of kind i*, by analogy with the terminology of the geometric theory of infinitely near points in the plane, where a point $O^{(i)}$, infinitely near the point $O^{(1)} \equiv (0, 0)$, is said to be of kind i, if it is in the $(i-1)$-th neighborhood of $O^{(1)}$. The identity of the two concepts will appear from the formal power series considerations of Part II. However, already at this stage, the analogy appears from the fact, that while it takes $i-1$ successive quadratic transformations to transform a point $O^{(i)}$ of kind i into a proper point (a point of kind 1), it takes as well $i-1$ successive quadratic transformations to transform a simple v-ideal $\mathfrak{P}_i$ of kind i into a simple v-ideal of kind 1, i. e. into a prime 0-dimensional ideal.

THEOREM 6. 1. *If $\{\mathfrak{P}_\alpha\}$ and $\{\bar{\mathfrak{P}}_\alpha\}$ are the sequences of simple v-ideals in $\mathfrak{O}$ belonging to valuations B and $\bar{B}$ respectively and if, for a given i, we have $\mathfrak{P}_i = \bar{\mathfrak{P}}_i$, then also $\mathfrak{P}_\alpha = \bar{\mathfrak{P}}_\alpha$ for any $\alpha < i$. In other words, the $i-1$ simple v-ideals which precede a given simple v-ideal $\mathfrak{P}_i$ of kind i in a given valuation B for which $\mathfrak{P}_i$ is a v-ideal, are uniquely determined by $\mathfrak{P}_i$, being independent of the valuation B.*

Proof by induction. The theorem is trivial for $i = 1$. Assume that the theorem is true for simple v-ideals of kind $i-1$, and apply the quadratic transformation T. We will have then in $\mathfrak{O}'$: $\mathfrak{P}'_{i-1} = \bar{\mathfrak{P}}'_{i-1}$, where $\mathfrak{P}'_{\alpha-1} = T(\mathfrak{P}_\alpha)$ and $\bar{\mathfrak{P}}'_{\alpha-1} = T(\bar{\mathfrak{P}}_\alpha)$. Hence, by our induction, $\mathfrak{P}'_\alpha = \bar{\mathfrak{P}}'_\alpha$, for $\alpha = 1, 2, \cdots, i-2$, whence $\mathfrak{P}_\alpha = \bar{\mathfrak{P}}_\alpha$ for $\alpha = 2, \cdots, i-1$, because $\mathfrak{P}_\alpha$ as well as $\bar{\mathfrak{P}}_\alpha$ is the contracted ideal of the ideal $x'^h\mathfrak{P}'_{\alpha-1}$ $(= x'^h\bar{\mathfrak{P}}'_{\alpha-1})$, where h is the smallest integer such that $x'^h\mathfrak{P}'_{\alpha-1}$ is an extended ideal of an ideal in $\mathfrak{O}$. Moreover, $\mathfrak{P}_1 = \mathfrak{P}'_1$, since $\mathfrak{P}_1$ and $\mathfrak{P}'_1$ are the prime ideals belonging to $\mathfrak{P}_i$ and $\mathfrak{P}'_i$ respectively q. e. d.

A much stronger theorem can be proved:

THEOREM 6.2. *Under the hypothesis $\mathfrak{P}_i = \bar{\mathfrak{P}}_i$ of Theorem* 6.1, *the set of v-ideals for B which precede $\mathfrak{P}_i$ (divisors of $\mathfrak{P}_i$) coincides with the set of v-ideals for $\bar{B}$ which precede $\bar{\mathfrak{P}}_i$.*

Proof by induction with respect to i. For $i = 1$ the theorem is trivial. Assume that the theorem is true for simple v-ideals of kind $i - 1$. Let $\mathfrak{q}_k$ be a v-ideal for B such that $\mathfrak{P}_i \equiv 0(\mathfrak{q}_k)$. If $\mathfrak{q}_k$ is simple, then it is also a v-ideal for $\bar{B}$, by the preceding theorem. If $\mathfrak{q}_k$ is composite, we know by Theorem 5.1 that it can be factored into simple v-ideals belonging to B. Since $\mathfrak{q}_k$ is a proper divisor of $\mathfrak{P}_i$, only factor $\mathfrak{P}_j$, $j < i$, can occur. Let

$$\mathfrak{q}_k = \mathfrak{P}_1^{\alpha_1}\mathfrak{P}_2^{\alpha_2}\cdots\mathfrak{P}_{i-1}^{\alpha_{i-1}}, \qquad \mathfrak{P}_1 = \mathfrak{p} = (x, y).$$

We consider separately two cases: (1) $\alpha_1 = 0$, (2) $\alpha_1 > 0$.

(1) First case: $\alpha_1 = 0$. We apply our quadratic transformation T. We find then

$$T(\mathfrak{q}_k) = \mathfrak{q}'_\sigma = \mathfrak{P}'^{\alpha_2}_1\cdots\mathfrak{P}'^{\alpha_{i-1}}_{i-2}. \tag{15}$$

Since the factor $\mathfrak{P}_1 (= \mathfrak{p})$ does not occur in the factorization of $\mathfrak{q}_k$, the system $\Omega(\mathfrak{q}_k)$ of subforms of $\mathfrak{q}_k$ is of dimension zero. Hence $\mathfrak{q}_k = T^{-1}(\mathfrak{q}'_\sigma)$. Since also $\mathfrak{P}_i = T^{-1}(\mathfrak{P}'_{i-1})$ and since $\mathfrak{P}_i \equiv 0(\mathfrak{q}_k)$, we have, by Theorem 4.4, $\mathfrak{P}'_{i-1} \equiv 0(\mathfrak{q}'_\sigma)$. Now $\mathfrak{P}'_{i-1}$ is a v-ideal in $\mathfrak{O}'$ for both B and $\bar{B}$. Hence, by our induction, $\mathfrak{q}'_\sigma$ is also a v-ideal for $\bar{B}$. But then also $\mathfrak{q}_k$ must be a v-ideal in $\mathfrak{O}$ for $\bar{B}$, since $\mathfrak{q}_k = T^{-1}(\mathfrak{q}'_\sigma)$.

(2) Second case: $\alpha_1 > 0$. We now use an induction with respect to k, i. e. we assume it has been already proved that all the v-ideals $\mathfrak{q}_j$ for B, $j < k$, are also v-ideals for $\bar{B}$. Since in the factorization (15) the factor $\mathfrak{p}$ occurs to the power α_1, it follows that the system $\Omega(\mathfrak{q}_k)$ of subforms of $\mathfrak{q}_k$ is of dimension α_1. Hence, by Theorem 3 (Corollary 3.2) we can write $\mathfrak{q}_k = \mathfrak{p}^{\alpha_1}\mathfrak{q}_t$, where $\mathfrak{q}_t$ is a v-ideal for B. Let $\mathfrak{p}^{\alpha_1-1}\mathfrak{q}_t \sim \mathfrak{q}_s$. We have $\mathfrak{q}_k \equiv 0(\mathfrak{p}\mathfrak{q}_s)$; on the other hand, since $v(\mathfrak{p}^{\alpha_1-1}\mathfrak{q}_t) = v(\mathfrak{q}_s)$, it follows that $v(\mathfrak{p}\mathfrak{q}_s) = v(\mathfrak{p}^{\alpha_1}\mathfrak{q}_t) = v(\mathfrak{q}_k)$, whence $\mathfrak{p}\mathfrak{q}_s \equiv 0(\mathfrak{q}_k)$. Consequently $\mathfrak{q}_k = \mathfrak{p}\mathfrak{q}_s$. We now consider the k-th v-ideal $\bar{\mathfrak{q}}_k$ for $\bar{B}$. We must suppose that the factor $\mathfrak{P}_1$ occurs also in the factorization of $\bar{\mathfrak{q}}_k$, since otherwise $\bar{\mathfrak{q}}_k$ would also be a v-ideal for B, by the preceding case $\alpha_1 = 0$, whence necessarily $\bar{\mathfrak{q}}_k = \mathfrak{q}_k$, since k is the length of both ideals $\mathfrak{q}_k$, $\bar{\mathfrak{q}}_k$. Let then $\bar{\mathfrak{q}}_k = \mathfrak{p}\bar{\mathfrak{q}}_\sigma$. By our induction, $\bar{\mathfrak{q}}_\sigma$ is a v-ideal for B, since $\sigma < k$; i. e. $\bar{\mathfrak{q}}_\sigma = \mathfrak{q}_\sigma$. Hence $\mathfrak{q}_k = \mathfrak{p}\mathfrak{q}_s$, $\bar{\mathfrak{q}}_k = \mathfrak{p}\mathfrak{q}_\sigma$. Since of the two ideals $\mathfrak{q}_s$, $\mathfrak{q}_\sigma$ one is a divisor of the other, the same is true of the ideals $\mathfrak{q}_k$ and $\bar{\mathfrak{q}}_k$. But these two ideals have the same length k, consequently $\mathfrak{q}_k = \bar{\mathfrak{q}}_k$, q. e. d.

Remark. The following example shows that a composite v-ideal does not determine the v-ideals preceding it. Let B be the valuation defined by the branch $y = x^{3/2}$ and let $\bar{B}$ be the valuation determined by the branch $y = x^{2/3}$. We have then $\mathfrak{q}_1 = \mathfrak{p} = (x, y)$, $\mathfrak{q}_2 = (y, x^2)$, $\mathfrak{q}_3 = \mathfrak{p}^2$, while $\bar{\mathfrak{q}}_1 = \mathfrak{p} = \mathfrak{q}_1$, $\bar{\mathfrak{q}}_2 = (x, y^2) \neq \mathfrak{q}_2$, $\bar{\mathfrak{q}}_3 = \mathfrak{p}^2 = \mathfrak{q}_3$.

7. A factorization theorem for v-ideals. We know from the preceding sections that any v-ideal $\mathfrak{A}$, belonging to a given valuation B, can be factored into simple v-ideals belonging to the same valuation B. The question arises as to the *unicity of this factorization.* The unicity of the factorization may be *a priori* intended in more than one way. In the first place, we *may fix some valuation B* to which $\mathfrak{A}$ belongs, and we may ask whether the factorization of $\mathfrak{A}$ into simple v-ideals *belonging to B* is unique. We may go a step further and ask whether the factorization of $\mathfrak{A}$, if unique for a given valuation B, *is independent of B.* Finally, we may formulate the unicity of factorization of $\mathfrak{A}$ in its strongest possible form and assert, that $\mathfrak{A}$ *can be factored in a unique manner into simple v-ideals,* where we allow *a priori* that the simple v-factors may belong to different valuations. It is this strongest form of the unicity theorem which we proceed to prove. It will, of course, follow from this theorem, that the simple v-factors are v-ideals for any valuation for which $\mathfrak{A}$ is a v-ideal.

We prove, however, a stronger theorem, from which the unique factorization of v-ideals into simple v-ideals will follow:

THEOREM 7.1. *Let $\mathfrak{A}_1, \mathfrak{A}_2, \cdots, \mathfrak{A}_k$ and $\bar{\mathfrak{A}}_1, \bar{\mathfrak{A}}_2, \cdots, \bar{\mathfrak{A}}_s$ be two sets of simple v-ideals belonging to valuations $B_1, B_2, \cdots, B_k$ and $\bar{B}_1, \bar{B}_2, \cdots, \bar{B}_s$ respectively. If*

$$\mathfrak{A}_1^{\alpha_1}\mathfrak{A}_2^{\alpha_2}\cdots\mathfrak{A}_k^{\alpha_k} = \bar{\mathfrak{A}}_1^{\bar{\alpha}_1}\bar{\mathfrak{A}}_2^{\bar{\alpha}_2}\cdots\bar{\mathfrak{A}}_s^{\bar{\alpha}_s}, \tag{16}$$

where the α_i's and $\bar{\alpha}_j$'s are positive integers, then necessarily $k = s$ and, for a proper arrangement of the indices, $\mathfrak{A}_i = \bar{\mathfrak{A}}_i$, $\alpha_i = \bar{\alpha}_i$.

This theorem is stronger, because a product of v-ideals, in particular the power product $\Pi\mathfrak{A}_i^{\alpha_i}$, is not necessarily a v-ideal.[7]

Proof. Let the simple v-ideals $\mathfrak{A}_i$ and $\bar{\mathfrak{A}}_j$ be of kind h_i and $\bar{h}_j$ respectively[8]; and let $m = \max.(h_1, \cdots, h_k, \bar{h}_1, \cdots, \bar{h}_s)$. The theorem is trivial in the case $m = 1$. In fact, in this case the ideals $\mathfrak{A}_i$ and $\bar{\mathfrak{A}}_j$ are prime zero-dimensional ideals, and the power products on both sides of (16) coincide with $[\mathfrak{A}_1^{\alpha_1}, \mathfrak{A}_2^{\alpha_2}, \cdots, \mathfrak{A}_k^{\alpha_k}]$ and $[\bar{\mathfrak{A}}_1^{\bar{\alpha}_1}, \bar{\mathfrak{A}}_2^{\bar{\alpha}_2}, \cdots, \bar{\mathfrak{A}}_s^{\bar{\alpha}_s}]$ respectively. The theorem follows in this case from the unicity of the decomposition of a zero-dimensional ideal into primary components. We may then prove our theorem by induction with respect to m.

---

[7] For instance, let $\mathfrak{A}_1 = (y, x^2)$, $\mathfrak{A}_2 = (x, y^2)$. Both $\mathfrak{A}_1$, $\mathfrak{A}_2$ are v-ideals, but $\mathfrak{A}_1\mathfrak{A}_2 = (xy, \mathfrak{p}^3)$ is not a v-ideal, because the system $\Omega(\mathfrak{A}_1\mathfrak{A}_2)$ of subforms is of dimension zero and its base xy is not the power of a linear form (see Theorem 3).

[8] We assume that all the ideals $\mathfrak{A}_i$, $\bar{\mathfrak{A}}_j$ are zero-dimensional (necessarily primary). The case of one-dimensional simple v-ideals is trivial, because any such ideal is prime,

The partial power products on both sides of (16) consisting of factors which belong to one and the same zero-dimensional ideal, must be equal to each other. Hence it is sufficient to prove the theorem for the case in which the ideals $\mathfrak{A}_i$, $\bar{\mathfrak{A}}_j$ all belong to one and the same prime ideal, say to $\mathfrak{p} = (x, y)$. The ideal $\mathfrak{p}$ consists then, for each of the given valuations B_i, $\bar{B}_j$, of all the polynomials whose value is > 0. We may assume that $v(x) = v(\mathfrak{p})$ in any of the valuations B_i, $\bar{B}_j$. Let us now apply the quadratic transformations $T: x' = x,\ y' = y/x$, and let $T(\mathfrak{A}_i) = \mathfrak{A}'_i$, $T(\bar{\mathfrak{A}}_j) = \bar{\mathfrak{A}}'_j$. If we have, in the valuation B_i, $v\left(\dfrac{y - c_i x}{x}\right) > 0$, then $\mathfrak{A}'_i$ will be a primary ideal in $\mathfrak{O}' = \mathfrak{k}[x', y']$ belonging to the ideal $(x', y' - c_i)$. A similar remark holds for $\bar{\mathfrak{A}}'_j$. At any rate, $\mathfrak{A}'_i$ will be a simple v-ideal of kind $h_i - 1$ and $\bar{\mathfrak{A}}'_j$ will be a simple v-ideal of kind $\bar{h}_j - 1$. We must remember, however, that the transform of a simple v-ideal of kind 1, i. e. of $\mathfrak{p}$, is the unit ideal $\mathfrak{O}'$. If then $\mathfrak{A}_1 = \bar{\mathfrak{A}}_1 = \mathfrak{p}$, where we allow now that one or both of the exponents α_1, $\bar{\alpha}_1$ may be zero, operating by T on (16) we get

$$\mathfrak{A}'^{\alpha_2}_2 \mathfrak{A}'^{\alpha_3}_3 \cdots \mathfrak{A}'^{\alpha_k}_k = \bar{\mathfrak{A}}'^{\bar{\alpha}_2}_2 \bar{\mathfrak{A}}'^{\bar{\alpha}_3}_3 \cdots \bar{\mathfrak{A}}'^{\bar{\alpha}_s}_s.$$

Since $\max.(h_i - 1, \bar{h}_j - 1) = m - 1$, we have by our induction, $k = s$, $\alpha_i = \bar{\alpha}_i$, $\mathfrak{A}'_i = \bar{\mathfrak{A}}'_i$ $(i > 1)$, and (16) becomes

$$\mathfrak{p}^{\alpha_1} \mathfrak{A}_2^{\alpha_2} \cdots \mathfrak{A}_k^{\alpha_k} = \mathfrak{p}^{\bar{\alpha}_1} \mathfrak{A}_2^{\alpha_2} \cdots \mathfrak{A}_k^{\alpha_k}.$$

Now α_1 is the dimension of the system $\Omega(\mathfrak{p}^{\alpha_1} \mathfrak{A}_2^{\alpha_2} \cdots \mathfrak{A}_k^{\alpha_k})$ and similarly $\bar{\alpha}_1$ is the dimension of $\Omega(\mathfrak{p}^{\bar{\alpha}_1} \mathfrak{A}_2^{\alpha_2} \cdots \mathfrak{A}_k^{\alpha_k})$. Hence $\alpha_1 = \bar{\alpha}_1$ and the theorem is proved.

PART II.

8. Algebraic and transcendental valuations.

In this part of the paper we shall use the apparatus of formal power series in order to derive further

---

and is therefore a principal ideal (f), where f is an irreducible polynomial. The one-dimensional factors and their exponents on both sides of (16) must be the same and may be deleted.

properties of valuation ideals in the ring of polynomials. The use of formal power series is clearly indicated by the fact that a zero-dimensional valuation of the field Σ of rational functions of x and y is essentially a local property of the field. It is known that any zero-dimensional valuation B of rank 1, in which x and y have positive values, can be obtained by the following construction: We put

$$x = A(t) = \alpha_1 t^{a_1} + \alpha_2 t^{a_2} + \cdots, \qquad y = B(t) = \beta_1 t^{b_1} + \beta_2 t^{b_2} + \cdots,$$

where the coefficients α and β belong to the underlying field $\mathfrak{k}$ and where the exponents a_i, b_i of each power series $A(t)$ and $Q(t)$ form a monotonic increasing sequence of positive real numbers. By substitution of these power series every element r of Σ takes a definite form

$$r = \gamma_1 t^{c_1} + \gamma_2 t^{c_2} \cdots (\gamma_1 \neq 0, c_1 \gtreqless 0),$$

and the valuation B is obtained by putting $v(r) = c_1$. We may eliminate formally t between $x = A(t)$ any $y = B(t)$ and we may thus define the valuation B by putting $y = P(x) = \delta_1 x^{d_1} + \delta_2 x^{d_2} + \cdots$, where the exponents are again increasing positive real numbers.

Now we may effect the substitution $x = A(t)$, $y = B(t)$ not only in any rational function r of x, y but also in any formal power series $\xi = \sum_{i,j} a_{ij} x^i y^j$, $i, j \geqq 0$, i, j-integers, and in any quotient of such formal power series. In this manner the valuation B defines a valuation B^* of the field Σ^* of meromorphic functions of x, y, and the valuation ring of B^* contains the ring $\mathfrak{O}^*$ of holomorphic functions ξ.

The special case in which the exponents a_i, b_i of the power series $A(t)$, $B(t)$ are integers, is the only one which is of interest in the classical theory of algebraic and analytic functions. The corresponding valuations B may be called *algebroid* or *analytic*, while non-analytic valuations may be referred to as *transcendental valuations*. For an algebroid valuation the power series $P(x)$ is an ordinary Puiseux series, i. e. the exponents $d_1, d_2, \cdots$ are rational numbers with fixed denominator:

$$d_i = m_i/n, \qquad (n, m_1, m_2, \cdots) = 1.$$

If the branch $y = P(x)$ is *algebraic*, i. e. belongs to an algebraic curve $f(x, y) = 0$, f-irreducible, then the valuation B is *algebraic* and is effectively of rank 2, being composed of the prime divisor defined by the prime ideal (f) and of a valuation of the field of rational functions on the curve $f(x, y) = 0$. In all cases, if B is algebroid, the induced valuation B^* of the field of meromorphic functions is of rank 2. If, namely, we denote by $P_0 = P, P_1, \cdots, P_{n-1}$

the n determinations of the power series $P(x)$, corresponding to the n determinations of $x^{1/n}$, then $\zeta = \prod_{i=0}^{n-1} (y - P_i)$ is an holomorphic function of x, y, i. e. an element of $\mathfrak{O}^*$ and is indecomposable in $\mathfrak{O}^*$. Given any element ξ of $\mathfrak{O}^*$, we have a unique decomposition $\xi = \zeta^{\rho}\xi_1$, $\xi_1 \not\equiv 0(\zeta)$. The substitution $y = P(x)$ does not annihilate ξ_1 and hence we find for ξ_1 a definite representation $\xi_1 = \gamma_1 x^{c_1} + \gamma_2 x^{c_2} + \cdots$, $\gamma_1 \neq 0$. We define B^* by putting $v(\xi) = (\rho, c_1)$.

It is evident that, conversely, any valuation B of Σ of rank 2 is algebraic, and if the values of x and y are positive, B can be defined by putting y equal to a Puiseux series in x, *provided that the divisor of which B is composed is of the first kind with respect to $\mathfrak{O}$.*

In the sequel we will have no occasion to use transcendental valuations. The results of Part I enable us, in fact, to prove the following theorem:

THEOREM 8.1. *Every valuation ideal in $\mathfrak{O}$ belongs to an algebroid (and even to an algebraic) valuation of Σ.*

Proof.[9] It is sufficient to prove this assertion for 0-dimensional (primary) v-ideals, because v-ideals possessing a 1-dimensional component can belong only to algebraic valuations. Let $\{\mathfrak{q}_i\}$ be the sequence of zero-dimensional v-ideals in $\mathfrak{O}$ belonging to a valuation B. We wish then to prove that given any ideal in the sequence, say $\mathfrak{q}_n$, there exists an algebraic valuation for which $\mathfrak{q}_n$ is a valuation ideal. The nature of our proof requires that a stronger assertion be established. We propose to prove that *there exists an algebraic valuation $\bar{B}$ such that in the Jordan sequence $\{\bar{\mathfrak{q}}_i\}$ of the 0-dimensional v-ideals belonging to $\bar{B}$, the first n ideals $\bar{\mathfrak{q}}_1, \bar{\mathfrak{q}}_2, \cdots, \bar{\mathfrak{q}}_n$ coincide with $\mathfrak{q}_1, \mathfrak{q}_2, \cdots, \mathfrak{q}_n$.* This we prove by induction with respect to n, assuming then that this assertion has been already established for $n - 1$, *for any choice of the generators x, y of Σ and for any valuation of Σ whose valuation ring contains $\mathfrak{k}[x, y]$.* Assuming, as usual, that x and y have positive values in B, we use the quadratic transformation $T: x' = x$, $y' = y/x$, getting the ring $\mathfrak{O} = \mathfrak{k}[x', y']$ and the sequence of v-ideals $\{\mathfrak{q}'_i\}$ in $\mathfrak{O}'$ belonging to B. By our induction, there exists an algebraic valuation $\bar{B}$ of Σ whose valuation ring contains $\mathfrak{O}'$ and such that the ideals $\mathfrak{q}'_1, \mathfrak{q}'_2, \cdots, \mathfrak{q}'_{n-1}$ are v-ideals belonging to $\bar{B}$. Let $\{\bar{\mathfrak{q}}_i\}$ be the sequence of v-ideals in $\mathfrak{O}$ belonging to $\bar{B}$. We now use the results of section 4. The $n - 1$ ideals $T^{-1}\mathfrak{q}'_i$, $i = 1, 2, \cdots, n-1$ must be members of both sequences $\{\mathfrak{q}_i\}$ and $\{\bar{\mathfrak{q}}_i\}$; here $T^{-1}\mathfrak{q}'_i$ is the contracted ideal of $x'^{h_i}\mathfrak{q}'_i$, where h_i is defined as the smallest integer such that $x'^{h_i}\mathfrak{q}'_i$ is an extended ideal of an ideal in $\mathfrak{O}$. Let $T^{-1}\mathfrak{q}'_i = \mathfrak{q}_{\alpha_i} = \bar{\mathfrak{q}}_{\alpha_i}$, $i = 1, 2, \cdots, n-1$ (the indices α_i are the same, since the index j of $\mathfrak{q}_j$ is its length). We also know that $\alpha_1 < \alpha_2 \cdots < \alpha_{n-1}$ and that $\alpha_{n-1} \geqq n$. Moreover, if $i \leqq \alpha_{n-1}$, then $\mathfrak{q}_i$ is necessarily of the form $\mathfrak{p}^\rho \mathfrak{q}_{\alpha_\sigma}$, $\sigma \leqq n-1$, and also $\bar{\mathfrak{q}}_i$ is of the form $\mathfrak{p}^{\bar{\rho}} \mathfrak{q}_{\alpha_{\bar{\sigma}}}$, $\bar{\sigma} \leqq n-1$. *We assert that* $\mathfrak{q}_i = \bar{\mathfrak{q}}_i$, $i = 1, 2, \cdots, \alpha_{n-1}$. Suppose that we know already that $\mathfrak{q}_j = \bar{\mathfrak{q}}_j$, for all $j < i \leqq \alpha_{n-1}$. If $\mathfrak{q}_i$ (or $\bar{\mathfrak{q}}_i$) coincides with one of the ideals $\mathfrak{q}_{\alpha_j}$, $j \leqq \alpha_{n-1}$, there is nothing to prove: we will have $\mathfrak{q}_i = \mathfrak{q}_{\alpha_\sigma} = \bar{\mathfrak{q}}_{\alpha_\sigma} = \bar{\mathfrak{q}}_i$. In the contrary case, we have $\mathfrak{q}_i = \mathfrak{p}^\rho \mathfrak{q}_{\alpha_\sigma}$, $\rho > 0$. It is not difficult to see that $\mathfrak{q}_i$ is then also of the form: $\mathfrak{q}_i = \mathfrak{p}\mathfrak{q}_j$, where necessarily $j < i$. In fact, let $\mathfrak{p}^{\rho-1}\mathfrak{q}_{\alpha_\sigma} \sim \mathfrak{q}_j$ (the equivalence being intended in the sense of the valuation B and $\mathfrak{q}_j$ being a v-ideal for B). Then $\mathfrak{p}^{\rho-1}\mathfrak{q}_{\alpha_\sigma} \equiv 0(\mathfrak{q}_j)$, whence $\mathfrak{q}_i \equiv 0(\mathfrak{p}\mathfrak{q}_j)$. On the other hand it is clear that $\mathfrak{q}_i \sim \mathfrak{p}\mathfrak{q}_j$, whence $\mathfrak{p}\mathfrak{q}_j \equiv 0(\mathfrak{q}_i)$. Hence $\mathfrak{q}_i = \mathfrak{p}\mathfrak{q}_j$. In a similar manner we find for $\bar{\mathfrak{q}}_i$ a representation of the form $\bar{\mathfrak{q}}_i = \mathfrak{p}\bar{\mathfrak{q}}_\mu$. Since $\mu < i$, we have $\bar{\mathfrak{q}}_\mu = \mathfrak{q}_\mu$, whence of the two ideals $\mathfrak{q}_j$ and $\bar{\mathfrak{q}}_\mu$ one is a divisor of the other ($\mathfrak{q}_j \equiv 0(\bar{\mathfrak{q}}_\mu)$ or $\bar{\mathfrak{q}}_\mu \equiv 0(\mathfrak{q}_j)$ according as $j \geqq \mu$ or $\mu \geqq j$). As a consequence it is also true that of the two ideals $\mathfrak{q}_i$ and $\bar{\mathfrak{q}}_i$, one is a divisor of the other. Now both $\mathfrak{q}_i$ and $\bar{\mathfrak{q}}_i$ have the same length i. Consequently $\mathfrak{q}_i = \bar{\mathfrak{q}}_i$, and this proves our assertion. Since the equality $\mathfrak{q}_i = \bar{\mathfrak{q}}_i$ holds for $i = 1, 2, \cdots, \alpha_{n-1}$ and since $\alpha_{n-1} \geqq n$, it follows that $\mathfrak{q}_1, \mathfrak{q}_2, \cdots, \mathfrak{q}_n$ belong as v-ideals to the algebraic valuation $\bar{B}$, and this proves our theorem.

---

[9] Note that the proof makes no use of the Theorem 5. 2.

Using Theorem 8.1 and the results of sections 5, 6, it is possible to give a very simple proof of the well-known fact that *every valuation B of Σ is the limit of algebraic valuations*. Let $\{\mathfrak{q}_i\}$ be the sequence of 0-dimensional v-ideals belonging to B. We may assume that the intersection of the ideals $\mathfrak{q}_i$ is the 0-ideal, since otherwise B itself is algebraic. In the proof of the Theorem 8.1 it has been shown that for any value of k there exists an algebraic valuation B_k of Σ, for which the ideals $\mathfrak{q}_1, \mathfrak{q}_2, \cdots, \mathfrak{q}_k$ are valuation ideals. Let r/s be any element of the valuation ring $\mathfrak{B}$ of B, $r, s \in \mathfrak{O}$. There will then exist an integer n such that $r \equiv 0(\mathfrak{q}_n)$, $s \not\equiv 0(\mathfrak{q}_{n+1})$. This integer will depend only on s. For all values of k such that $k \geqq n+1$, the ideals $\mathfrak{q}_n$ and $\mathfrak{q}_{n+1}$ will also be v-ideals for B_k, and hence for all such values of k r/s will also belong to the valuation ring $\mathfrak{B}_k$ of B_k. As a consequence, we have $\lim\limits_{k\to\infty} (\mathfrak{B}_k \cap \mathfrak{B}) = \mathfrak{B}$, i. e. B can be regarded as the limit of the valuation B_k.

Using the characteristic property of the Jordan sequence $\{\mathfrak{q}_i\}$ of v-ideals established in section 2, and the properties of simple ideals given in sections 5 and 6, we may go a step further and gain an insight into the manner in which transcendental valuations are constructed. Given a simple v-ideal $\mathfrak{P}_{k+1}$ we know that it determines uniquely the sequence of simple v-ideals $\mathfrak{P}_1, \mathfrak{P}_2, \cdots, \mathfrak{P}_k$ which precede $\mathfrak{P}_{k+1}$ in any valuation to which $\mathfrak{P}_{k+1}$ belongs. We ask now the following question: given $\mathfrak{P}_k$ (and hence given the entire sequence $\mathfrak{P}_1, \mathfrak{P}_2, \cdots, \mathfrak{P}_k$)

in how many ways is it possible to choose $\mathfrak{P}_{k+1}$? To answer this question, we apply $k-1$ successive quadratic transformations, getting a ring $\bar{\mathfrak{O}}$ of polynomials of variables X, Y, in which to $\mathfrak{P}_k$ there corresponds a simple v-ideal $\bar{\mathfrak{P}}_1$ of kind one, i. e. $\bar{\mathfrak{P}}_1$ is prime and 0-dimensional, say $\bar{\mathfrak{P}}_1 = (X, Y)$. Any maximal subideal of $\bar{\mathfrak{P}}_1$ is of the form $(aX + bY, \bar{\mathfrak{P}}_1^2)$, where a, b are in $\mathfrak{k}$ and are not both zero. It is obvious that any such maximal subideal of $\bar{\mathfrak{P}}_1$ belongs to some valuation (for instance, to any valuation defined by putting $X = bt + \cdots$, $Y = -at + \cdots$) and is moreover a simple v-ideal. These maximal subideals are in $(1, 1)$ correspondence with the ratio $z = a/b$, i. e. with the places of the purely transcendental field $\mathfrak{k}(z)$. Going back to our original ring $\mathfrak{O}$, we see that the set of simple v-ideals $\mathfrak{P}_{k+1}$ of kind $k+1$ such that $\mathfrak{P}_1, \mathfrak{P}_2, \cdots, \mathfrak{P}_k, \mathfrak{P}_{k+1}$ belong to one and the same valuation, is in $(1, 1)$ correspondence with the set consisting of the elements of the underlying field $\mathfrak{k}$ and of the symbol ∞. Starting with $\mathfrak{P}_1$ we can then construct, in infinitely many ways, an infinite sequence $\mathfrak{P}_1, \mathfrak{P}_2, \cdots, \mathfrak{P}_k, \cdots$ of simple v-ideals, such that, for any k, the ideals $\mathfrak{P}_1, \mathfrak{P}_2, \cdots, \mathfrak{P}_k$ belong to some valuation B_k, which we may suppose to be algebraic. We assert that *the infinite sequence* $\{\mathfrak{P}_k\}$ *defines a valuation* B *of* Σ. Obviously then $B = \operatorname{Lim} B_k$. To see this, we first observe, that by Theorem 6. 2, the infinite sequence $\{\mathfrak{P}_k\}$ determines uniquely an infinite Jordan sequence $\{\mathfrak{q}_j\}$ which contains the sequence $\{\mathfrak{P}_k\}$ and which has the property that the elements of the sequence which precede a given $\mathfrak{P}_i$ are v-ideals for all the valuations B_k, $k \geqq i$. It follows immediately that the congruence $\mathfrak{q}_{m+1}\mathfrak{q}_n : \mathfrak{q}_m \equiv 0(\mathfrak{q}_{n+1})$ holds true for any two ideals $\mathfrak{q}_m$, $\mathfrak{q}_n$ of the sequence $\{\mathfrak{q}_j\}$, since the ideals which occur in this congruence are v-ideals belonging to B_k, when k is sufficiently large. As a consequence the sequence effectively defines a valuation of Σ, *q. e. d.*

9. Valuation ideals in the ring of holomorphic functions.[10] It has been pointed out in the preceding section that any 0-dimensional valuation B of the field $\Sigma = \mathfrak{k}(x, y)$ in which x and y have positive values, defines a valuation B^* of the field Σ^* of meromorphic functions of x, y, whose valuation ring contains the ring $\mathfrak{O}^* = \mathfrak{k}\{x, y\}$ of holomorphic functions of x, y. We have then also valuation ideals in $\mathfrak{O}^*$ belonging to B^*. It is clear that the prime ideal defined by B^* in $\mathfrak{O}^*$ is the 0-dimensional ideal $\mathfrak{p}^* = (x, y)$, i. e. the extended ideal of the ideal $\mathfrak{p} = (x, y)$ in $\mathfrak{O}$. Let $\{\mathfrak{q}_i\}$ be the sequence of 0-dimensional v-ideals in $\mathfrak{O}$ belonging to B, and let $\mathfrak{q}^*_i = \mathfrak{O}^*\mathfrak{q}_i$ be the extended ideal of $\mathfrak{q}_i$ in $\mathfrak{O}^*$. It is clear that $v(q^*_{i+1}) > v(\mathfrak{q}^*_i)$ in B^*, since $v(\mathfrak{q}^*_i) = v(\mathfrak{q}_i)$. Let $f_1, f_2, \cdots, f_k$ be a base of $\mathfrak{q}_i$ and let f be an element of $\mathfrak{q}_i$ not in $\mathfrak{q}_{i+1}$. Since $\mathfrak{q}_i/\mathfrak{q}_{i+1} \cong \mathfrak{k}$, we have $f_i \equiv c_i f(\mathfrak{q}_{i+1})$, $c_i \in \mathfrak{k}$. If then $\xi = \xi_1 f_1 + \cdots + \xi_k f_k$ is any element of $\mathfrak{q}^*_i$ ($\xi_j \in \mathfrak{O}^*$), we have $\xi \equiv f \cdot (c_1\xi_1 + \cdots + c_k\xi_k)(\mathfrak{q}^*_{i+1})$. Now $c_1\xi_1 + \cdots + c_k\xi_k \equiv c(\mathfrak{p}^*)$, where $c \in \mathfrak{k}$, and

---

[10] Results of this and of the following sections will be later applied toward the proof of Theorem 5. 2. The properties of simple v-ideals derived in Part I and based on Theorem 5. 2 are therefore not to be used until a proof of this theorem has been given (Corollary 11. 2). We may, however, use Theorem 8. 1 (see footnote on p. 175).

$$(f)\mathfrak{p}^* \equiv 0(\mathfrak{q}^*_i\mathfrak{p}^*) \equiv 0(\mathfrak{O}^*\mathfrak{q}_i\mathfrak{p}) \equiv 0(\mathfrak{q}^*_{i+1}),$$

since $\mathfrak{q}_i\mathfrak{p} \equiv 0(\mathfrak{q}_{i+1})$. Hence $\xi \equiv cf(\mathfrak{q}^*_{i+1})$, and this shows that $\mathfrak{q}^*_i/\mathfrak{q}^*_{i+1} \simeq \mathfrak{k}$, i. e. $\mathfrak{q}^*_{i+1}$ is a maximal subideal of $\mathfrak{q}^*_i$. Hence the sequence $\{\mathfrak{q}^*_i\}$ is a Jordan sequence in $\mathfrak{O}^*$, and since $v(\mathfrak{q}^*_{i+1}) > v(\mathfrak{q}^*_i)$, the sequence $\{\mathfrak{q}^*_i\}$ is the sequence of v-ideals in $\mathfrak{O}^*$ belonging to B^*, i. e. *the zero-dimensional v-ideals in $\mathfrak{O}^*$ belonging to B^* are the extended ideals of the zero-dimensional v-ideals in $\mathfrak{O}$ belonging to B.* It is of course evident that $\mathfrak{q}_i$ is the contracted ideal of $\mathfrak{q}^*_i$.

The theorems of sections 3, 4,[11] the notions of simple and composite ideals introduced in section 5 and Theorem 5. 1 carry over without modification to ideals in $\mathfrak{O}^*$. In order to see this, a perusal of the proof is not at all necessary. It is sufficient to take into account quite generally the nature of the relationship between the polynomial ideals in $\mathfrak{O}$ and the power series ideals in $\mathfrak{O}^*$. There is a (1, 1) correspondence between primary 0-dimensional ideal in $\mathfrak{O}$ belonging to the prime ideal $\mathfrak{p} = (x, y)$ and the 0-dimensional ideals in $\mathfrak{O}^*$. The correspondence is such that to an ideal $\mathfrak{q}$ in $\mathfrak{O}$ there corresponds its extended ideal $\mathfrak{q}^*$ in $\mathfrak{O}^*$ and $\mathfrak{q}$ is the contracted ideal of $\mathfrak{q}^*$. In fact, let $\mathfrak{q}^*$ be any 0-dimensional ideal in $\mathfrak{O}^*$. If ρ is the exponent of $\mathfrak{q}^*$, the $\mathfrak{q}^*$ possesses a base consisting of the power products $x^i y^j$, $i + j = \rho$, and of a set of polynomials F_α of degree $< \rho$. Hence $\mathfrak{q}^* = (F_\alpha, x^i y^j)$ and therefore $\mathfrak{q}^*$ is the extended ideal of a zero-dimensional primary polynomial ideal $\mathfrak{q}$ belonging to $\mathfrak{p}$ ($= (x, y)$). Let $\mathfrak{q}^* = (f_1, f_2, \cdots, f_k)$, where $f_1, f_2, \cdots, f_k$ is a base of $\mathfrak{q}$, and let $\mathfrak{q}'$ be the contracted ideal in $\mathfrak{O}$ of $\mathfrak{q}^*$. If F is any polynomial in $\mathfrak{q}'$, $F = \sum_{i=1}^{k} \xi_i(x, y) f_i$, $\xi_i \in \mathfrak{O}^*$, and if we denote by $A_i^{(n)}$ the partial sum of the terms of degree $\leqq n$ in the power series $\xi_i(x, y)$, then the polynomial $F - \Sigma A_i^{(n)} f_i$ does contain terms of degree $< n + 1$, whence $F \equiv \Sigma A_i^{(n)} f_i(\mathfrak{p}^{n+1})$. Now, if n is sufficiently high, then $\mathfrak{p}^{n+1} \equiv 0(\mathfrak{q})$, whence $F \equiv 0(\mathfrak{q})$. Hence $\mathfrak{q}' \equiv 0(\mathfrak{q})$, and since $\mathfrak{q}'$ is the contracted ideal of $\mathfrak{q}^*$, it follows $\mathfrak{q}' = \mathfrak{q}$. Thus every zero-dimensional ideal $\mathfrak{q}^*$ in $\mathfrak{O}^*$ is the extended ideal of one and only one primary ideal $\mathfrak{q}$ in $\mathfrak{O}^*$, and $\mathfrak{q}$ is the contracted ideal of $\mathfrak{q}^*$. In particular, $\mathfrak{q}^*$ is simple or composite, according as $\mathfrak{q}$ is simple or composite.

---

[11] In the case of formal power series the quadratic transformation $x' = x$, $y' = y/x$ leads from the ring $k\{x, y\}$ of formal power series in x and y to the larger ring of $k\{x', y'\}$ of formal power series in x' and y'.

10. The notion of a general element of an ideal of formal power series. The ring $\mathfrak{O}^*$ of holomorphic functions of x and y contains only one prime 0-dimensional prime ideal, namely the ideal $\mathfrak{p}^* = (x, y)$, and every 0-dimensional ideal $\mathfrak{A}$ in $\mathfrak{O}^*$ is necessarily primary and belongs to $\mathfrak{p}^*$. Since any ideal in $\mathfrak{O}^*$ has a finite base, given a 0-dimensional ideal $\mathfrak{A}$, it belongs to a finite exponent ρ, i. e. $\mathfrak{p}^{*\rho} \equiv 0(\mathfrak{A})$. In other words, $\mathfrak{A}$ contains all the formal power series which contain terms of lowest degree $\geqq \rho$. Since $\mathfrak{A}$ is at any rate a linear $\mathfrak{k}$-module, it follows that the condition in order that an element $\xi = \sum\limits_{i,j \geqq 0} a_{ij}, x^i y^j$ of $\mathfrak{O}^*$ belong to $\mathfrak{A}$ is expressed by linear homogeneous relations between the coefficients a_{ij}, $i + j \leqq \rho$. With every linear relation

$$(17) \qquad \sum_{i,j < \rho} c_{ij} a_{ij} = 0, \qquad c_{ij} \, \varepsilon \, \mathfrak{k},$$

satisfied by all the elements ξ of $\mathfrak{A}$, we associate the function

$$(17') \qquad E = \sum c_{ij} x^{-i} y^{-j}.$$

The set of all the functions E obtained in this manner is called the "inverse system" $\mathfrak{A}^{-1}$ of the ideal $\mathfrak{A}$ (Macaulay,[5,6] Lasker[3]). If $E_1, E_2, \cdots, E_r$ belong to $\mathfrak{A}^{-1}$, then also $\alpha_1 E_1 + \cdots + \alpha_r E_r$ is in $\mathfrak{A}^{-1}$, $\alpha_1, \cdots, \alpha_r \, \varepsilon \, \mathfrak{k}$. From the fact that if $\xi = \Sigma a_{ij} x^i y^j$ belongs to the ideal $\mathfrak{A}$, also

$$x\xi = \Sigma a_{ij} x^{i+1} y^j \quad \text{and} \quad y\xi = \Sigma a_{ij} x^i y^{j+1}$$

belong to $\mathfrak{A}$, it follows immediately that if $E = c_{ij} x^{-i} y^{-j}$ is an element of the inverse system, then also the following relations are true for any element $\xi = \Sigma a_{ij} x^i y^j$ in $\mathfrak{A}$:

$$\sum c_{i+1,j} a_{ij} = 0, \qquad \sum c_{i,j+1} a_{ij} = 0.$$

This shows that the inverse system $\mathfrak{A}^{-1}$ *becomes an $\mathfrak{O}^*$-module, provided that we define multiplication as follows:*

$$x^\alpha y^\beta \sum_{i,j \geqq 0} c_{ij} x^{-i} y^{-j} = \sum_{\substack{i \geqq \alpha \\ j \geqq \beta}} c_{ij} x^{-i+\alpha} y^{-j+\beta}, \qquad \alpha, \beta \geqq 0.$$

In fact, with this definition of multiplication by power products $x^\alpha y^\beta$, the above equations $\Sigma c_{i+1,j}a_{ij}=0$, $\Sigma c_{i,j+1}a_{ij}=0$ signify that if E is in $\mathfrak{A}^{-1}$, also xE and yE are in $\mathfrak{A}^{-1}$. Multiplication of E by any formal power series in $\mathfrak{O}^*$ is then to be defined formally by the requirement of the distributive law of multiplication. It is then not difficult to see that $\mathfrak{A}^{-1}$ *consists of those and only those functions E which have the property that $\xi E=0$ for any element ξ in $\mathfrak{A}$*. It is evident that if $\mathfrak{A}$ and $\mathfrak{B}$ are ideals in $\mathfrak{O}^*$, then $\mathfrak{A}^{-1}=\mathfrak{B}^{-1}$ implies $\mathfrak{A}=\mathfrak{B}$. Moreover, it is not difficult to show that given any set $\mathfrak{C}^{-1}$ of functions E which is an $\mathfrak{O}^*$-module and for which there exists at least one element $\xi\neq 0$ in $\mathfrak{O}^*$ such that $\xi E=0$ for any E in $\mathfrak{C}^{-1}$, then the set of all elements ξ satisfying this condition forms an ideal $\mathfrak{A}$, and $\mathfrak{C}^{-1}$ is the inverse system of $\mathfrak{A}$.

We now come to the definition of a concept which shall be very useful in the sequel. Let $\tau=t_{ij}x^iy^j$ be a formal power series whose coefficients t_{ij} belong to an extension field $\mathfrak{K}$ of $\mathfrak{k}$. We shall say that τ is a *variable element of an ideal* $\mathfrak{A}$, if the coefficients t_{ij} do not all belong already to $\mathfrak{k}$ (so that at least one of the t_{ij}'s is transcendental with respect to $\mathfrak{k}$) and if they satisfy (in $\mathfrak{K}$) all the linear relations (17) (with a_{ij} replaced by t_{ij}), which are satisfied by the coefficients of all the actual power series belonging to $\mathfrak{A}$. In other words, we require that $\tau E=0$ for any function E in $\mathfrak{A}^{-1}$. We shall say that a variable element τ of $\mathfrak{A}$ is a *general element of* $\mathfrak{A}$, if the t_{ij}'s do not satisfy any other algebraic relation, algebraically independent of the above linear relations. Finally, we shall say that a variable element τ of $\mathfrak{A}$ is a *quasi-general* element of $\mathfrak{A}$, if the t_{ij}'s do not satisfy *linear relations* other than those which are satisfied by the coefficients of a general element of $\mathfrak{A}$.

In a similar manner we define the notions of a variable element, of a general element and of a quasi-general element of the inverse system $\mathfrak{A}^{-1}$. The condition that a function $E=\Sigma c_{ij}x^{-i}y^{-j}$ belong to $\mathfrak{A}^{-1}$ is expressed by a certain set (δ) of homogeneous linear equations between the c_{ij}'s. These equations are obtained by expressing the fact that $E\xi_i=0$, $i=1,2,\cdots,k$, where $\xi_1,\xi_2,\cdots,\xi_k$ is a base of the ideal $\mathfrak{A}$. A function $E=\Sigma e_{ij}x^{-i}y^{-j}$, where the e_{ij} are elements of an extension field $\mathfrak{k}(e_{ij})$ of $\mathfrak{k}$, shall be called a *variable element of* $\mathfrak{A}^{-1}$, if the coefficients e_{ij} satisfy all the relations of the set (δ), i. e. if $E\xi=0$ for any ξ in $\mathfrak{A}$. If the coefficients e_{ij} do not satisfy algebraic (*or* linear) relations algebraically independent of the linear relations (δ), then E shall be called a *general* (*or quasi-general*) element of $\mathfrak{A}^{-1}$.

Let $E=\Sigma e_{ij}x^{-i}y^{-j}$ and $\tau=\Sigma t_{ij}x^iy^j$ be variable elements of $\mathfrak{A}^{-1}$ and of $\mathfrak{A}$ respectively, where we assume that the e_{ij} and t_{ij} are elements of one and the same field $\mathfrak{K}$.

We have $\xi E=0$ for any element $\xi=\Sigma a_{ij}x^iy^j$ of $\mathfrak{A}$. Since the a_{ij}'s are arbitrary elements of underlying field $\mathfrak{k}$ subject to the only condition of satisfying a finite set of linear relations $\Sigma c_{ij}a_{ij} = 0$, where $\Sigma c_{ij}x^{-i}y^{-j}$ is an element of $\mathfrak{A}^{-1}$, it follows that the relation $\xi E = 0$ remains true if we regard the a_{ij} as indeterminates connected by the linear relations $\Sigma c_{ij}a_{ij} = 0$. But then the relation $\xi E = 0$ holds also after the specialization $a_{ij} \to t_{ij}$, since the t_{ij} also satisfy all the relations $\Sigma c_{ij}t_{ij} = 0$. *We conclude that* $\tau E = 0$. The following assertion is now easily derived:

If E *is a quasi-general element of* $\mathfrak{A}^{-1}$, *then* $\tau E = 0$ *implies that* τ *is a variable element of* $\mathfrak{A}$, *and if* τ *is a quasi-general element of* $\mathfrak{A}$ *then* $\tau E = 0$ *implies that* E *is a variable element of* $\mathfrak{A}^{-1}$. Proof straightforward.

Let $\mathfrak{A}', \mathfrak{A}''$ be two (distinct or coincident) ideals in $\mathfrak{D}^*$ and let $\tau' = \Sigma t'_{ij}x^iy^j$, $\tau'' = \Sigma t''_{ij}x^iy^j$ be variable elements of $\mathfrak{A}'$ and $\mathfrak{A}''$ respectively. Let $\mathfrak{K}' = \mathfrak{k}(t'_{ij})$ and $\mathfrak{K}'' = \mathfrak{k}(t''_{ij})$. We consider a pure transcendental base $\{u\}$ of $\mathfrak{K}''$ over $\mathfrak{k}$ and we adjoin the elements of this base to the field $\mathfrak{K}'$, *adjunction to be regarded as a pure transcendental extension of* $\mathfrak{K}'$. We obtain in this manner the field $\mathfrak{K}'(\{u\})$ having $\mathfrak{k}(\{u\})$ as subfield, and we then adjoin to $\mathfrak{K}'(\{u\})$ all the elements of $\mathfrak{K}''$ which are algebraic with respect to $\mathfrak{k}(\{u\})$. In the field $\mathfrak{K} = \mathfrak{k}(t'_{ij}, t''_{ij})$ obtained in this manner, any algebraic relation between the t'_{ij} and the t''_{ij} must be a consequence of algebraic relations between the t'_{ij}'s alone in $\mathfrak{K}'$ and the t''_{ij} alone in $\mathfrak{K}''$. Now that the t'_{ij} and the t''_{ij} have been properly imbedded in a common field, we form the product $\tau'\tau'' = \tau = \Sigma t_{ij}x^iy^j$, $t_{ij} \varepsilon \mathfrak{K}$, and we refer to τ as *the direct product of the variable elements* τ', τ'' *of* $\mathfrak{A}'$ *and* $\mathfrak{A}''$.

THEOREM 10. 1. *The direct product* $\tau = \tau'\tau''$ *of variable elements of two ideals* $\mathfrak{A}'$, $\mathfrak{A}''$ *is a variable element of the product* $\mathfrak{A}'\mathfrak{A}''$ *of the two ideals. If* τ', τ'' *are general or quasi-general elements of* $\mathfrak{A}'$ *and* $\mathfrak{A}''$ *respectively, then* τ *is a quasi-general* (*not necessarily general*) *element of* $\mathfrak{A}'\mathfrak{A}''$.

Proof. Let $\mathfrak{A}'\mathfrak{A}'' = \mathfrak{B}$. It is well known (and is of immediate verification) that $\mathfrak{B}^{-1} = \mathfrak{A}'^{-1} : \mathfrak{A}''$, i. e. $\mathfrak{B}^{-1}$ consists of all functions E such that $E\mathfrak{A}'' \varepsilon \mathfrak{A}'^{-1}$. Let then $E_0 = \Sigma c_{ij}x^{-i}y^{-j}$ be an element in $\mathfrak{B}^{-1}$ and let $\tau''E_0 = \Sigma \sigma_{ij}x^{-i}y^{-j}$, $\sigma_{ij} \varepsilon \mathfrak{K}$. *We assert that* $\tau''E_0$ *is a variable element of* $\mathfrak{A}'^{-1}$. In fact, consider any element $\xi'' = \Sigma a''_{ij}x^iy^j$ of $\mathfrak{A}''$. Since $E_0\mathfrak{A}'' \varepsilon \mathfrak{A}'^{-1}$, the product $\xi''E_0 = \Sigma \sigma_{ij}{}^{(0)}x^{-i}y^{-j}$ is an element of the inverse system $\mathfrak{A}'^{-1}$. Hence the coefficients $\sigma_{ij}{}^{(0)}$ must satisfy the linear relations (δ') which are satisfied by the coefficients of the general elements of $\mathfrak{A}'^{-1}$. The coefficients $\sigma_{ij}{}^{(0)}$ being linear forms in the coefficients a''_{ij} of ξ'', we get by substitution into the equations (δ') a set of linear homogeneous relations between the a''_{ij}. Since these relations hold true for the coefficients a''_{ij} of any element ξ'' of $\mathfrak{A}''$, it follows that they are also satisfied by the coefficients t''_{ij} of the variable element τ'' of $\mathfrak{A}''$. But then it follows that the coefficients σ_{ij} of $\tau''E_0$ must satisfy the linear equations (δ''), whence $\tau''E_0$ is a variable element of $\mathfrak{A}'^{-1}$. Since τ' is a variable element of $\mathfrak{A}'$, it follows by a previously proved result, that $\tau'\tau''E_0 = 0$, i. e. $\tau E_0 = 0$. Since the relation $\tau E_0 = 0$ holds for any element E_0 of $\mathfrak{B}^{-1}$, it follows that τ is a variable element of $\mathfrak{B}$, and this proves the first part of the theorem.

Now suppose that τ' and τ'' are general (or quasi-general) elements of the ideals $\mathfrak{A}'$ and $\mathfrak{A}''$ respectively. To prove that in this case τ is a quasi-general element of $\mathfrak{B}$ we have to show that if $\Sigma c_{ij}t_{ij} = 0$ is any linear homogeneous relation between the t_{ij}, then $E = \Sigma c_{ij}x^{-i}y^{-j}$ belongs to the inverse system $\mathfrak{B}^{-1}$. Now the t_{ij} are bilinear forms in the two sets of coefficients t'_{ij} and t''_{ij}. Substituting these bilinear forms we get $\Sigma c_{ij}t_{ij} = H(t'_{ij}, t''_{ij})$, where H is a bilinear form in the t'_{ij} and the t''_{ij}. The relation $H(t'_{ij}, t''_{ij}) = 0$ must be a consequence of the *linear relations* between the t'_{ij} and the t''_{ij} separately, since τ is the direct product of τ' and τ''. Now τ' is a quasi-general element of $\mathfrak{A}'$. As a consequence, any *linear* relation between the t'_{ij} arises from a function E' in $\mathfrak{A}'^{-1}$ and is therefore not destroyed if each t'_{ij} is replaced by $t'_{i-1,j}$ or by $t'_{i,j-1}$ (multiplication of E' by x or by y respectively). Hence the relation $H(t'_{ij}, t''_{ij}) = 0$ implies the relations $H(t'_{i-1,j}, t''_{ij}) = 0$ and $H(t'_{i,j-1}, t''_{ij}) = 0$. It is clear that these two relations correspond to the relations $\Sigma c_{ij}t_{i-1,j} = 0$ and $\Sigma c_{ij}t_{i,j-1} = 0$, which therefore must be true relations between the t_{ij}. As a consequence the functions $E = c_{ij}x^{-i}y^{-j}$ corresponding to the various linear relations $\Sigma c_{ij}t_{ij} = 0$ between the t_{ij}, form an $\mathfrak{O}^*$-module. From this it follows immediately that $E\tau = 0$, for any E in this module. Now $E\tau = E\tau'\tau'' = 0$ implies that $E\tau'$ is a variable element of $\mathfrak{A}''^{-1}$, since τ'' is a quasi-general element of $\mathfrak{A}''$. The condition of belonging to $\mathfrak{A}''^{-1}$ is expressed by a certain set (δ'') of linear equations, and thus the coefficients of $E\tau'$ must satisfy these equations. Since τ' is a quasi-general element of $\mathfrak{A}'$, these linear equations (δ'') must be satisfied by $E\xi'$, ξ'—an arbitrary element of $\mathfrak{A}'$, and hence $E\mathfrak{A}' \varepsilon \mathfrak{A}''^{-1}$. As a consequence $E \varepsilon \mathfrak{A}''^{-1} : \mathfrak{A}' = \mathfrak{B}^{-1}$, which proves our theorem.[12]

---

[12] That τ need not be the general element of $\mathfrak{A}'\mathfrak{A}''$ is shown by the following example. Let $\mathfrak{p} = (x, y)$, $\mathfrak{q} = (x^2, y^2)$, so that $\mathfrak{p}\mathfrak{q} = \mathfrak{p}^3$. The general element τ' of $\mathfrak{p}$ is $t'_{10}x + t'_{01}y + t'_{20}x^2 + \cdots$, where the t'_{ij}'s are indeterminates. Similarly

$$\tau'' = t''_{20}x^2 + t''_{02}y^2 + t''_{30}x^3 + \cdots$$

is the general element of $\mathfrak{q}$. Now

$$\tau = \tau'\tau'' = t_{30}x^3 + t_{21}x^2 + t_{12}xy^2 + t_{03}y^3 + \cdots,$$

where the t_{ij} satisfy *one non-linear* relation $t_{30}t_{03} = t_{21}t_{12}$.

The preceding theorem implies that if an ideal $\mathfrak{B}$ is composite, then a suitable quasi-general element τ of $\mathfrak{B}$ (namely the direct product of the general elements of the factors of $\mathfrak{B}$) is reducible in the algebraic closure of the field of the coefficients t_{ij} of τ. We are interested in the question of the extent to which the above considerations can be inverted. *What can be said about an ideal $\mathfrak{B}$, if its general element τ is reducible in the algebraic closure of the coefficients of τ?* That the ideal $\mathfrak{B}$ need not in general be composite, is illustrated by the example $\mathfrak{B} = (x^2, y^2)$. The ideal $\mathfrak{B}$ is simple, but its general element

$$\tau = t_{20}x^2 + t_{02}y^2 + t_{30}x^3 + \cdots \quad (t_{00} = t_{10} = t_{01} = t_{11} = 0)$$

is reducible, since

$$\tau = (\sqrt{t_{20}}\,x + \sqrt{-t_{02}}\,y + \cdots)(\sqrt{t_{20}}\,x - \sqrt{-t_{02}}\,y + \cdots).$$

However, in the special case of valuation ideals we can prove the following theorem:

THEOREM 10.2. *If $\mathfrak{A}$ is a valuation ideal and if the general element t of $\mathfrak{A}$ is reducible, $t = t_1 t_2 \cdots t_k$ (in the algebraic closure of its coefficients), then $\mathfrak{A}$ is composite; t is the direct product of its irreducible factors t_i, each irreducible factor t_i is the general element of a valuation ideal $\mathfrak{A}_i$, and $\mathfrak{A} = \mathfrak{A}_1 \mathfrak{A}_2 \cdots \mathfrak{A}_k$.*

Proof. Let $t = \Sigma t_{ij} x^i y^j$ and let $t = t_1 t_2 \cdots t_k$ be the factorization of t into irreducible factors $t_i = t_i(x, y)$, belonging to the ring of formal power series of x, y with coefficients in the algebraic closure of $\mathfrak{k}(\cdots t_{ij} \cdots)$. We assume of course that no t_i is a unit, i. e. that t_i does not contain a term independent of x and y. The valuation B to which $\mathfrak{A}$ belongs can be assumed to be algebraic. Substituting into t and into t_i the Puiseux expansion $y = P(x)$ which determines the valuation B, we are able to attach to these elements definite values $v(t)$, $v(t_i)$. It is clear that $v(t) = v(\mathfrak{A})$. Let $v(t_i) = \alpha_i$, whence $v(t) = \alpha_1 + \alpha_2 + \cdots + \alpha_k$. There exist elements in $\mathfrak{k}\{x, y\}$ whose value in B equals α_i: they can be obtained by specializing the coefficients of the formal power series t_i in such a manner as not to annihilate the coefficient of the leading term x^{α_i} of $t_i(x, P(x))$. As a consequence there exists a v-ideal $\mathfrak{A}_i$ for B such that $v(\mathfrak{A}_i) = \alpha_i$. Since $\alpha_i > 0$, no $\mathfrak{A}_i$ is the unit ideal. Since $v(t_i) = v(\mathfrak{A}_i)$, it follows that t_i is a variable element of $\mathfrak{A}_i$, and hence $t = \Pi t_i$

is a variable element of the ideal $\mathfrak{A}_1\mathfrak{A}_2\cdots\mathfrak{A}_k$. But since t is a general element of $\mathfrak{A}$, necessarily $\mathfrak{A}\equiv 0(\mathfrak{A}_1\mathfrak{A}_2\cdots\mathfrak{A}_k)$. On the other hand we have

$$v(\mathfrak{A})=v(t)=\Sigma v(t_i)=\Sigma v(\mathfrak{A}_i)=v(\Pi\mathfrak{A}_i).$$

Hence $\Pi\mathfrak{A}_i\equiv 0(\mathfrak{A})$, and consequently $\mathfrak{A}=\mathfrak{A}_1\mathfrak{A}_2\cdots\mathfrak{A}_k$.

Let τ_i be the general element of $\mathfrak{A}_i$ and let us consider the *direct* product $\tau=\tau_1\tau_2\cdots\tau_k$. By Theorem 10.1, τ is a variable element of $\mathfrak{A}$. On the other hand, any algebraic relation between the coefficients of the power series τ leads also to a true relation between the coefficients of t, since t is obtained from τ by the specialization $\tau_i\to t_i$. Consequently τ *must be a general element of* $\mathfrak{A}$. It can be then identified with t, and it is thus seen that in the original factorization $t=\Pi t_i$, each t_i is a general element of $\mathfrak{A}_i$ and that the product is a direct product (unicity of factorization of t into irreducible factors). The theorem is proved.

COROLLARY. *The general element of a simple v-ideal is absolutely irreducible* (i.e. irreducible in $\mathfrak{K}\{x,y\}$, $\mathfrak{K}$ being any extension field of $\mathfrak{k}$).

11. The characterization of simple *v*-ideals. Let

$$(18)\qquad y=y_1=\sum_{i=1}^{k}c_ix^{\alpha_i/\nu}+\sum_{j=0}^{\infty}t_jx^{(\alpha_{k+1}+j)/\nu},\qquad 0<\alpha_1<\alpha_2<\cdots$$

be a Puiseux series, *in which we assume that* $(\nu,\alpha_1,\alpha_2,\cdots,\alpha_g)=1$, $g\leqq k+1$, *and that the first* k *coefficients* $c_1,\cdots,c_k$ *are in the underlying field* $\mathfrak{k}$, *while the remaining coefficients are indeterminates.*

THEOREM 11.1. *Given a formal power series* $\xi(x,y)=\Sigma a_{ij}x^iy^j$ *with indeterminate coefficients* a_{ij}, *there exists a set of linear forms* $F_m(a_{ij})$, $G_n(a_{ij})$ *which have the following properties:* (1) *the relations* $F_m(a_{ij}^{(0)})=0\,(a_{ij}^{(0)}\,\varepsilon\,\mathfrak{k})$ *give necessary conditions that the equation* $\xi_0(x,y)=\Sigma a_{ij}^{(0)}x^iy^j=0$ *admit a uniformization of type* (18), *the* t_j*'s being replaced by special values* $t_j^{(0)}$ *in* $\mathfrak{k}$; (2) *the relations* $F_m(a_{ij}^{(0)})=0$ *and the inequalities* $G_n(a_{ij}^{(0)})\neq 0$ *give necessary and sufficient condition in order that* $\xi_0(x,y)=0$ *admit the above uniformization and that* $\xi_0(x,y)$ *be an irreducible element of* $\mathfrak{k}\{x,y\}$. *The set of all elements* ξ *in* $\mathfrak{k}\{x,y\}$ *whose coefficients* a_{ij} *satisfy the relations* $F_m=0$, *is a simple ideal.*

Proof. The theorem is true for $g=0$, in which case $\nu=1$. In fact, let

$\bar{y} = \sum_{i=1}^{k} c_i x^{a_i}$ ($\bar{y} = 0$, if $k = 0$). If the equation $\xi(x, y) = 0$ is uniformizable by the expansion $y = y_1$, then we must have $\xi(x, y_1) = 0$, and this shows that $\xi(x, \bar{y})$ is divisible by $x^{a_{k+1}}$. This condition is expressed by a certain set of linear homogeneous equations $F_m(a_{ij}) = 0$ between the coefficients of $\xi(x, y)$. Should moreover $\xi(x, y)$ be irreducible, then we must have

$$\xi(x, y) = (y - y_1)\eta(x, y),$$

where $\eta(x, y)$ is a unit. As a consequence, the coefficient a_{01} must be different from 0. Conversely, if $F_m(a_{ij}) = 0$ and $a_{01} \neq 0$, then $\xi(x, y)$ is divisible by $y - y_1$ and is irreducible in $\mathfrak{k}\{x, y\}$.

We assume that the theorem is true for $g - 1$. Let $\nu = \nu_1 n$, $\alpha_1 = \nu_1 \alpha'_1$, $(n, \alpha'_1) = 1$. We put

$$x = \bar{x}^n, \quad y = \bar{x}^{\alpha'_1}(c_1 + \bar{y}). \tag{19}$$

If $\xi(x, y) = \Sigma a_{ij} x^i y^j$ can be uniformized by the expansion $y = y_1$, then ξ must be divisible by the product $\prod_{i=1}^{\nu} (y - y_1^{(i)})$, where $y_1 = y_1^{(1)}, \cdots, y_1^{(\nu)}$ are the conjugates of y_1. From this it follows immediately that ξ cannot contain terms $x^i y^j$ in which $\nu i + \alpha_1 j < \nu\alpha_1$, whence

$$a_{ij} = 0, \text{ for all } i, j, \text{ such that } \nu i + \alpha_1 j < \nu\alpha_1. \tag{20}$$

Substituting (19) we find, in view of (20),

$$\xi(x, y) = \bar{x}^{\nu\alpha'_1}\bar{\xi}(\bar{x}, \bar{y}) = \bar{x}^{\nu\alpha'_1}\Sigma\bar{a}_{ij}\bar{x}^i\bar{y}^j, \tag{20'}$$

while the equation of the branch (18) becomes

$$\bar{y} = \bar{y}_1 = \sum_{i=2}^{k} c_i \bar{x}^{(a_i - a_1)/\nu_1} + \sum_{j=0}^{\infty} t_j \bar{x}^{(a_{k+1}+j-a_1)/\nu_1}. \tag{21}$$

The equation $\bar{\xi}(\bar{x}, \bar{y}) = 0$ admits a uniformization by means of an expansion $\bar{y} = \bar{y}_1$ of type (21). Since $\alpha_2 > \alpha_1$, there can be no constant term in $\bar{\xi}(\bar{x}, \bar{y})$. This constant term arises from the terms $a_{ij}x^i y^j$ of $\xi(x, y)$ in which $\nu i + \alpha_1 j = \nu\alpha_1$, i. e. $ni + \alpha'_1 j = n\alpha'_1\nu_1$, and is therefore equal to

$$a_{\nu_1\alpha'_1, 0} + a_{(\nu_1-1)\alpha'_1, n} c_1^{n} + \cdots + a_{0, n\nu_1} c_1^{n\nu_1}.$$

Consequently

$$a_{\nu_1\alpha'_1, 0} + a_{(\nu_1-1)\alpha'_1, n} c_1^{n} + \cdots + a_{0, n\nu_1} c_1^{n\nu_1} = 0. \tag{22}$$

It is clear that $(\alpha_2 - \alpha_1, \alpha_3 - \alpha_1, \cdots, \alpha_g - \alpha_1, \nu_1) = 1$, hence we are in the case $g-1$. By our induction, we have a set of linear forms $\bar{F}(\bar{a}_{ij})$ and $\bar{G}(\bar{a}_{ij})$ which satisfy the assertion of the theorem. Now the coefficients $\bar{a}_{ij}$ of $\bar{\xi}(\bar{x},\bar{y})$ are linear homogeneous forms in the coefficients a_{ij} of $\xi(x,y)$. Let $\{F_m\}$ be the set of forms in the a_{ij} consisting of the forms $\bar{F}(\bar{a}_{ij})$, expressed in terms of the a_{ij}, and of the left-hand members of the equations (20) and (22). Let moreover, $\{G_n\}$ be the set of forms in the a_{ij} consisting of the forms $\bar{G}(\bar{a}_{ij})$ (expressed in terms of the a_{ij}) and of the form $a_{0,\nu}$. We assert that the forms $\{F_m\}$ and $\{G_n\}$ satisfy the assertion of our theorem. In the first place, by the definition of the forms F_m, the equations $F_m=0$ must be satisfied if $\xi(x,y)$ can be uniformized by an expansion $y=y_1$ of the type (18). Assume that the coefficients a_{ij} satisfy the equations $F_m=0$ and the inequalities $G_n\neq 0$. The validity of the equations (20) implies that the substitution (19) introduces in $\xi(x,y)$ a factor $\bar{x}^{\lambda}$, where $\lambda \geqq \nu\alpha'_1$. Hence in (20′) the factor $\bar{\xi}(\bar{x},\bar{y})$ contains no negative powers of $\bar{x}$ or of $\bar{y}$. The coefficients $\bar{a}_{ij}$ of $\bar{\xi}$ satisfy, by hypothesis, the equations $\bar{F}(\bar{a}_{ij})=0$ and the inequalities $\bar{G}(\bar{a}_{ij})\neq 0$, relative to the branch (21). Hence the equation $\bar{\xi}(\bar{x},\bar{y})=0$ can be uniformized by an expansion $\bar{y}=\bar{y}_1$ of type (21), and consequently also the equation $\xi(x,y)=0$ can be uniformized by an expansion $y=y_1$ of type (18). The power series $\xi(x,y)$ must be divisible by the irreducible element $\prod_{j=1}^{\nu}(y-y_1^{(j)})$. Let $\xi=\eta(x,y)\cdot\prod_{j=1}^{\nu}(y-y_1^{(j)})$. But, by assumption $a_{0,\nu}\neq 0$, i. e. $\xi(x,y)$ contains a term in y^{ν}. Consequently $\eta(x,y)$ must contain a constant term $\neq 0$, since the coefficient of any term y^j, $j<\nu$, in the product $\prod_{j=1}^{\nu}(y-y_1^{(j)})$, is divisible by x. Hence η is a unit, and ξ is irreducible.

Conversely, if $\xi=0$ admits a uniformization $y=y_1$ of type (18) and if ξ is irreducible, we will have $\xi=\eta(x,y)\prod_{j=1}^{\nu}(y-y_1^{(j)})$, where η is a unit. This implies in the first place $a_{0\nu}\neq 0$, and it also implies, as was pointed out before, that the equations $F_m=0$ hold true. The inequalities $G_n(a_{ij})\neq 0$ arising from the inequalities $\bar{G}(\bar{a}_{ij})\neq 0$ must also be satisfied, since the hypothesis that ξ is irreducible in $\mathfrak{k}\{x,y\}$ implies that $\bar{\xi}(\bar{x},\bar{y})$ is irreducible in $\mathfrak{k}\{\bar{x},\bar{y}\}$.

It remains to prove that the elements $\xi(x,y)=\Sigma a_{ij}x^iy^j$ whose coefficients satisfy the linear relations $F_m=0$ form a simple ideal $\mathfrak{P}$. Let us regard the coefficients a_{ij} as elements of the field defined by the equations $F_m(a_{ij})=0$. The inequalities $G_n\neq 0$ are then satisfied, and $\xi=0$ admits a uniformization $y=y_1$ of type (18). As a consequence also $x\xi=0$ and $y\xi=0$ admit the uniformization $y=y_1$, and hence the coefficients of the two power series $x\xi$

and $y\xi$ must satisfy the relations $F_m = 0$. This implies that the relations $F_m(a_{i-1,j}) = 0$, $F_m(a_{i,j-1}) = 0$ are consequences of the relations $F_m(a_{ij}) = 0$, i. e. if ξ is any element in $\mathfrak{P}$, then $x\xi \varepsilon \mathfrak{P}$ and $y\xi \varepsilon \mathfrak{P}$. Hence $\mathfrak{P}$ is an ideal. To prove that $\mathfrak{P}$ is a simple ideal, we observe, that if $\mathfrak{P}$ was a composite ideal, then, by Theorem 10. 1, a suitable quasi-general element t of $\mathfrak{P}$ would be reducible in the algebraic closure of the field of the coefficients of t. Now t, a quasi-general element of $\mathfrak{P}$, has the property that its coefficients satisfy no linear relations other than those which hold for the coefficients of the general element of $\mathfrak{P}$, i. e. only the relations $F_m = 0$. Hence the coefficients of t certainly satisfy the inequalities $G_n \neq 0$, and therefore t could not be reducible, in contradiction with our assumption that $\mathfrak{P}$ is composite, q. e. d.

In order to apply the above theorem, we begin with some preliminary remarks. Let $t = \Sigma t_{ij}x^iy^j$ be the general element of a valuation ideal $\mathfrak{q}$, and let us assume that t is absolutely irreducible (this is certainly the case if $\mathfrak{q}$ is a simple v-ideal, see Theorem 10. 2, Corollary). We will have then $t = \epsilon . \prod_{i=1}^{\nu} (y - y_i)$, where ϵ is a unit and where $y_1, y_2, \cdots, y_\nu$ are the ν determinations of a Puiseux series $y_1 = \sum_{i=1}^{\infty} t_i x^{i/\nu}$, the t_i's being algebraic functions of the t_{ij}. Some of the coefficients t_i may be constants, i. e. elements in $\mathfrak{k}$. Let $t_i = c_i$, $i = 1, 2, \cdots, h$, $c_i \varepsilon \mathfrak{k}$, while t_{h+1} is the first coefficient which is transcendental with respect to $\mathfrak{k}$. The ideal $\mathfrak{q}$ is a valuation ideal for some algebraic valuation B, defined by an expansion $y = \eta = \sum_{i=1}^{\infty} d_i x^{i/\mu}$, $d_i \varepsilon \mathfrak{k}$, and the value of t, i. e. the evaluation of $\mathfrak{q}$, is the exponent of the term of lowest degree in $t(x, \eta) = \epsilon . \prod_{i=1}^{\nu} (y - y_i)$. It is clear that the term of lowest degree in $\eta - y_i$ is the same as the term of lowest degree in $\eta - \Sigma c_i x^{i/\nu} - t_{h+1}x^{(h+1)/\nu}$. It follows that the value of t is not altered if we regard the coefficients $t_{h+1}, t_{h+2}, \cdots$ in the expansion y_1 as entirely independent indeterminates. Now if $\bar{t}(x, y)$ denotes the *direct* product $\epsilon . \prod_{i=1}^{\nu} (y - y_i)$, in which $t_{h+1}, t_{h+2}, \cdots$ are regarded as indeterminates and ϵ is a unit with indeterminate coefficients (i. e. ϵ is the general element of the unit ideal), then $\bar{t}$ is a variable element of $\mathfrak{q}$, since $v(\bar{t}) = v(\mathfrak{q})$, and on the other hand $\bar{t}$ is at least as general a power series as t (i. e. t is a specialization of $\bar{t}$). Since t is the general element of $\mathfrak{q}$, it follows that t can be identified with $\bar{t}$, and hence the coefficients $t_{h+1}, t_{h+2}, \cdots$ in the original Puiseux series y_1 are indeed algebraically independent with respect to $\mathfrak{k}$, and can be regarded as indeterminates. Changing slightly our notation and putting into evidence the coefficients c_i which are different from zero, we re-write the series y_1 as follows:

$$(23) \qquad y_1 = \sum_{i=1}^{k} c_i x^{\alpha_i/\nu} + \sum_{j=0}^{\infty} t_j x^{(\alpha_{k+1}+j)/\nu}, \qquad \begin{matrix} 0 < \alpha_1 < \alpha_2 < \cdots, \\ c_1 c_2 \cdots c_k \neq 0, \end{matrix}$$

where $t_0, t_1, t_2,$ are indeterminates. Let δ be the highest common divisor of $\alpha_1, \alpha_2, \cdots, \alpha_{k+1}, \nu$. *We assert that if $\delta > 1$, then $t = \epsilon \prod_{i=1}^{\nu} (y - y_i)$ cannot be the general element of an ideal.* We shall show, namely, that if $\delta > 1$, then the coefficients t_{ij} of $t = \Sigma t_{ij} x^i y^j$ satisfy non-linear relations (which are not consequences of linear relations). Consider another sample of the series (23),

$$y'_1 = \sum_{i=1}^{k} c_i x^{\alpha_i/\nu} + \sum_{j=0}^{\infty} t'_j x^{(\alpha_{k+1}+j)/\nu},$$

where the t'_j are new indeterminates, and let $t' = \sum t'_{ij} x^i y^j = \epsilon' \prod_{i=1}^{\nu} (y - y'_i)$, where ϵ' is a unit with indeterminate coefficients. Also t' is a general element of $\mathfrak{q}$. Assume that the coefficients t_{ij} (and hence also the coefficients t'_{ij}) satisfy only linear homogeneous relations. Then it is clear that $t + t'$ is also a general element of $\mathfrak{q}$, whence $t + t' = \bar{\epsilon} \prod_{i=1}^{\nu} (y - \bar{y}_i)$, where $\bar{\epsilon}$ is a unit and

$$\bar{y}_1 = \sum_{i=1}^{k} c_i x^{\alpha_i/\nu} + \sum_{j=0}^{\infty} \tau_j x^{(\alpha_{k+1}+j)\nu}.$$

The substitution $y = \bar{y}_1$ must annihilate $t + t'$, i. e. $\epsilon \prod_{i=1}^{\nu} (y - y_i) + \epsilon' \prod_{i=1}^{\nu} (y - y'_i)$. We proceed to express the fact that the coefficients of the first two terms of $\epsilon \prod_{i=1}^{\nu} (\bar{y}_1 - y_i) + \epsilon' \prod_{i=1}^{\nu} (\bar{y}_1 - y'_i)$ vanish. If ω denotes a primitive ν-th root of unity, then

$$y_i = \sum_{j=1}^{k} c_j (\omega^i x^{1/\nu})^{\alpha_j} + t_0 (\omega^i x^{1/\nu})^{\alpha_{k+1}} + \sum_{j=1}^{\infty} t_j (\omega^i x^{1/\nu})^{\alpha_{k+1}+j},$$

whence

$$(24) \qquad \begin{aligned} \bar{y}_1 - y_i = {} & \sum_{j=1}^{k} c_j (1 - \omega^{i\alpha_j}) x^{\alpha_j/\nu} + (\tau_0 - t_0 \omega^{i\alpha_{k+1}}) x^{\alpha_{k+1}/\nu} \\ & + (\tau_1 - t_1 \omega^{i(\alpha_{k+1}+1)}) x^{(\alpha_{k+1}+1)/\nu} + \cdots. \end{aligned}$$

Let δ' be the h. c. d. of $\alpha_1, \alpha_2, \cdots, \alpha_k, \nu$, whence $\delta \equiv 0(\delta')$. We denote by $\Pi'(\bar{y}_1 - y_i)$ the product of the δ' factors $\bar{y}_1 - y_i$ for which $i\delta' \equiv 0(\nu)$, and by $\Pi''(\bar{y}_1 - y_i)$ the product of the remaining factors $\bar{y}_1 - y_i$, so that

$$\Pi'(\bar{y}_1 - y_i)\, \Pi''(\bar{y}_1 - y_i) = \prod_{i=1}^{\nu} (\bar{y}_1 - y_i).$$

If $i\delta' \not\equiv 0(\nu)$, the differences $1 - \omega^{i\alpha_j}$, $j = 1, 2, \cdots, k$, cannot all vanish.

Let σ be the smallest value of j, such that $1-\omega^{i\alpha_\sigma}\neq 0$. Then, by (24), $c_\sigma(1-\omega^{i\alpha_\sigma})x^{\alpha_\sigma/\nu}$, $\sigma\leqq k$, is the term of smallest degree in $\bar{y}_1-y_i$, while the exponent of the next term will be greater than $\alpha_\sigma/\nu+1/\nu$, since, for $j=1,2,\cdots,k$, $\alpha_{j+1}-\alpha_j\equiv 0(\delta)$, whence $\alpha_{j+1}-\alpha_j\geqq\delta>1$. It follows that

$$(25)\qquad \Pi''(\bar{y}_1-y_i)=dx^{\lambda}+d_1x^{\lambda_1}+\cdots,\ 0\neq d\,\varepsilon\,\mathfrak{k},\ \lambda_1>\lambda+1/\nu.$$

Similarly, replacing the t_j by the t'_j, we will have

$$(26)\qquad \Pi''(\bar{y}_1-y'_i)=dx^{\lambda}-d'_1x^{\lambda_1}+\cdots.$$

We now consider the product $\Pi'(\bar{y}_1-y_i)$. Here i assumes the values ν/δ', $2\nu/\delta'$, $\cdots$, $\delta'\nu/\delta'$, whence

$$(27)\qquad \begin{aligned}\bar{y}_1-y_i=\bar{y}-y_{j\nu/\delta'}&=(\tau_0-t_0\omega_1{}^{j\alpha_{k+1}})x^{\alpha_{k+1}/\nu}\\&\quad+(\tau_1-t_1\omega_1{}^{j(\alpha_{k+1}+1)})x^{(\alpha_{k+1}+1)/\nu}+\cdots,\ \omega_1=\omega^{\nu/\delta'},\\&\qquad\qquad (j=1,2,\cdots,\delta'),\end{aligned}$$

where ω_1 is a primitive root of unity of exponent δ'. Taking into account that $(\delta',\alpha_{k+1})=\delta>1$ and letting $\delta'=\delta h$, we find from (27),

$$\Pi'(\bar{y}_1-y_i)=(\tau_0{}^h-t_0{}^h)^{\delta}x^{\delta'\alpha_{k+1}/\nu}+\delta h(\tau_0{}^h-t_0{}^h)^{\delta-1}\tau_0{}^{h-1}\tau_1x^{(\delta'\alpha_{k+1}+1)/\nu}+\cdots.$$

It follows, by (25), that

$$\prod_{i=1}^{\nu}(\bar{y}_1-y_i)=d(\tau_0{}^h-t_0{}^h)^{\delta}x^{\lambda+\delta'\alpha_{k+1}/\nu}+d\delta h(\tau_0{}^h-t_0{}^h)^{\delta-1}\tau_0{}^{h-1}\tau_1x^{(\delta'\alpha_{k+1}+1)/\nu+\lambda}+\cdots.$$

The first two terms of $\prod_{i=1}^{\nu}(\bar{y}_1-y'_i)$ are obtained from the above by replacing t_0 by t'_0. Hence we must have

$$\epsilon_{00}(\tau_0{}^h-t_0{}^h)^{\delta}+\epsilon'_{00}(\tau_0{}^h-t'_0{}^h)^{\delta}=0,\quad \epsilon_{00}(\tau_0{}^h-t_0{}^h)^{\delta-1}+\epsilon'_{00}(\tau_0{}^h-t'_0{}^h)^{\delta-1}=0,$$

where ϵ_{00} and ϵ'_{00} are the constant terms in ϵ and ϵ' respectively. These equations imply the relations $\tau_0{}^h=t_0{}^h=t'_0{}^h$, in contradiction with the algebraic independence of the t_i and the t'_i. This proves our assertion.

This result shows that the general element $t=\epsilon\prod_{i=1}^{\nu}(y-y_i)$ of the valuation ideal $\mathfrak{q}$ is exactly of the type considered in the proof of Theorem

11.1, and hence $\mathfrak{q}$ belongs to the class of ideals $\mathfrak{P}$ defined in that theorem. This holds true for any valuation ideal $\mathfrak{q}$ whose general element is absolutely irreducible, in particular for any simple valuation ideal. But it has been shown that the ideals $\mathfrak{P}$ are simple ideals. Hence, *a valuation ideal whose general element is irreducible is simple.* The converse has already been proved before (Theorem 10.2, Corollary). Reassuming and recalling Theorem 10.2, we have the following theorem:

THEOREM 11.2. *A valuation ideal $\mathfrak{P}$ is simple if and only if its general element $t = \Sigma t_{ij}x^iy^j$ is absolutely irreducible. The general element t of $\mathfrak{P}$ is the direct product $t = \epsilon \cdot \prod_{i=1}^{\nu} (y - y_i)$, where ϵ is a unit (with indeterminate coefficients) and $y_1 = \sum_{i=1}^{k} c_i x^{\alpha_i/\nu} + \sum_{j=0}^{\infty} t_j x^{(\alpha_{k+1}+j)\nu}$, $(\alpha_1, \alpha_2, \cdots, \alpha_{k+1}, \nu) = 1$, the t_j being indeterminates. Given any valuation ideal $\mathfrak{A}$, the factorization of the general element of $\mathfrak{A}$ into irreducible factors yields a factorization of $\mathfrak{A}$ into simple v-ideals, according to the scheme indicated in Theorem* 10.2.

We are now in position to prove the basic Theorem 5.2 of Part I:

COROLLARY 11.2 (Theorem 5.2). *The transform of a simple v-ideal by a quadratic transformation is a simple v-ideal.*

Proof. It is immaterial whether the theorem is proved for polynomial ideals or for power series ideals, in view of the relationship between these ideals described in section 9. Let $\mathfrak{P}$ be a simple 0-dimensional ideal v-ideal in $\mathfrak{O}^* = \mathfrak{k}\{x, y\}$, belonging to a valuation B^*, and let $x'^{\nu}\mathfrak{P}'$ be the extended ideal of $\mathfrak{P}$ in $\mathfrak{O}'^* = \mathfrak{k}\{x', y'\}$, where $x = x'$, $y = y'x'$ (we assume as usual that $v(y) > v(x)$) and where ν is the integer such that $\mathfrak{P} \equiv 0(\mathfrak{p}^{*\nu})$, $\mathfrak{P} \not\equiv 0(\mathfrak{p}^{*\nu+1})$, $\mathfrak{p}^* = (x, y)$. Here $\mathfrak{P}'$ is necessarily either zero-dimensional or the unit ideal. Let $t(x, y) = \Sigma t_{ij}x^iy^j$ be the general element of $\mathfrak{P}$. By the definition of the integer ν, we must have $t_{ij} = 0$ for all i, j such that $i + j < \nu$, and $t_{ij} \neq 0$ for some i, j such that $i + j = \nu$. Hence $t(x, y) = t(x', x'y') = x'^{\nu}\tau(x', y')$, where $\tau(x', y') = \Sigma \tau_{ij}x'^iy'^j$ is a formal power series in x', y'. Here $\tau_{ij} = t_{i+\nu-j,j}$, if $i + \nu \geqq j$, and $\tau_{ij} = 0$, if $i + \nu < j$. In our special case the general element t of $\mathfrak{P}$ is of the form indicated in Theorem 11.2. Hence

$$\tau = \epsilon(x, y) \prod_{i=1}^{\nu} (y' - y'_i),$$

where

$$y'_1 = \sum_{i=1}^{k} c_i x'^{(\alpha_i-\nu)/\nu} + \sum_{j=0}^{\infty} t_j x'^{(\alpha_{k+1}+j-\nu)/\nu}.$$

If then $\epsilon'(x', y')$ denote a general unit of $\mathfrak{D}'^*$ (i. e. a unit with indeterminate coefficients), then the direct product $\epsilon'\tau$ is the general element of a simple ideal $\bar{\mathfrak{P}}'$. We have $v(\bar{\mathfrak{P}}') = v(\tau)$ and

$$v(\tau) + v(x'^{\nu}) = v(t) = v(\mathfrak{P}),$$

whence $v(\bar{\mathfrak{P}}') = v(\mathfrak{P}')$ and $\bar{\mathfrak{P}}' \equiv 0(\mathfrak{P}')$.

On the other hand, any element $t^0 = \Sigma t_{ij}{}^0 x^i y^j$ of $\mathfrak{P}$ is obtained from the general element t by a specialization $t_{ij} \rightarrow t_{ij}{}^0$. Hence, if

$$t^0(x, y) = x'^{\nu}\tau^0(x', y') = x'^{\nu}\Sigma\tau_{ij}{}^{(0)}x'^i y'^j,$$

then τ^0 is obtained from $\epsilon'\tau$ by the specialization $\epsilon' \rightarrow 1$, $\tau_{ij} \rightarrow \tau_{ij}{}^0$, and therefore τ^0 is an element of the ideal $\bar{\mathfrak{P}}'$. Since $x'^{\nu}\mathfrak{P}'$ is the extended ideal of $\mathfrak{P}$, a finite number of elements such as τ^0 form a base of $\mathfrak{P}'$. Hence $\mathfrak{P}' \equiv 0(\bar{\mathfrak{P}}')$, and consequently $\mathfrak{P}' = \bar{\mathfrak{P}}'$, q. e. d.

Remark 1. In view of the unicity of the factorization of a v-ideal $\mathfrak{A}$ into simple v-ideals (Theorem 7. 1), the second part of Theorem 11. 2 implies that the factorization of the general element of $\mathfrak{A}$ into irreducible factors yields *the* factorization of $\mathfrak{A}$ into simple v-ideals.

Remark 2. Concerning the characterization of simple v-ideals given in Theorem 11. 2, it is not difficult to show that, conversely, *an ideal* $\mathfrak{P}$ *whose general element* t *is of the type described in Theorem* 11. 2, *is a* v*-ideal* (necessarily simple). For the proof we assume $\alpha_1 > \nu$ and we consider the transform $\mathfrak{P}'$ of $\mathfrak{P}$ by the quadratic transformation $x' = x$, $y' = y/x$: $\mathfrak{D}'^*\mathfrak{P} = x'^{\nu}\mathfrak{P}'$. Using the notation of the proof above, we find as above the congruence $\mathfrak{P}' \equiv 0(\bar{\mathfrak{P}}')$. The preceding proof of the congruence $\bar{\mathfrak{P}}' \equiv 0(\mathfrak{P}')$ was based upon the fact that $\mathfrak{P}'$ was a v-ideal. But we may proceed without making use of this property of $\mathfrak{P}'$. Let $t'(x', y')$ be the general element of $\mathfrak{P}'$, $t' = \Sigma t'_{ij}x'^i y'^j$, and let $\Sigma c_{ij}t'_{ij} = 0$ be a true linear relation between the t'_{ij}. If $t^{(0)} = \Sigma t_{ij}{}^{(c)}x^i y^j$ is any element in $\mathfrak{P}$, and if $t^{(0)} = x'^{\nu}\tau_0$, $\tau_0 = \Sigma\tau_{ij}{}^{(0)}x'^i y'^j$, then τ_0 is an element of $\mathfrak{P}'$. Hence we must have

$$\Sigma\, c_{ij}\tau_{ij}{}^{(0)} = \sum_{i+\nu\geqq j} c_{ij}t^{(0)}{}_{i+\nu-j,j} = 0.$$

This relation holds for any element $t^{(0)}$ in $\mathfrak{P}$, consequently we must have $\sum_{i+\nu \geqq j} c_{ij} t_{i+\nu-j,j} = 0$, whence $\sum c_{ij}\tau_{ij} = 0$. This last relation shows that τ *is a variable element of* $\mathfrak{P}'$. But then, in view of Theorem 10.1, also $\epsilon'\tau$ is a variable element of $\mathfrak{P}'$, since ϵ' is the general element of the unit ideal. Consequently $\bar{\mathfrak{P}}' \equiv 0(\mathfrak{P}')$, whence again $\mathfrak{P}' = \bar{\mathfrak{P}}'$.

Consider the valuation B defined by the branch

$$y = \zeta = \sum_{i=1}^{k} c_i x^{\alpha_i/\nu} + \sum_{j=0}^{\infty} d_j x^{(\alpha_{k+1}+j)/\nu},$$

where $d_0, d_1, \cdots$ are arbitrary constants (in $\mathfrak{k}$) and $d_0 \neq 0$. Let the value of t in this valuation be λ_0/ν, λ_0—an integer. The integer $\lambda_0 = \lambda(\mathfrak{P})$ is independent of the particular constants d_j and is uniquely determined by the ideal $\mathfrak{P}$ and by the auxiliary condition $\alpha_1 \geqq \nu$ which prevents a special position of the axis $x = 0$. If the variables x', y' are used, then the valuation B is defined by the branch

$$y' = \zeta' = \sum_{i=1}^{k} c_i x'^{(\alpha_i-\nu)/\nu} + \sum_{j=0}^{\infty} d_j x^{(\alpha_{k+1}+j-\nu)/\nu}.$$

We have $v(\mathfrak{P}') = v(\mathfrak{P}) - v(x^\nu)$, whence $v(\mathfrak{P}') = (\lambda_0 - \nu^2)/\nu$. If $\alpha_1 - \nu \geqq \nu$, then $\lambda(\mathfrak{P}') = \lambda_0 - \nu^2 < \lambda(\mathfrak{P})$. If $\alpha_1 - \nu < \nu$, then the rôles of the variables x', y' must be interchanged, and putting $\alpha_1 - \nu = \nu'$ we find $v(\mathfrak{P}') = [(\lambda_0 - \nu^2)/\nu] \cdot (\nu/\nu')$, whence again $\lambda(\mathfrak{P}') = \lambda_0 - \nu^2 < \lambda(\mathfrak{P})$. The inequality $\lambda(\mathfrak{P}') < \lambda(\mathfrak{P})$ leads to a complete induction with respect to $\lambda(\mathfrak{P})$. If $\lambda(\mathfrak{P}) = 1$, then $\nu = 1$, since evidently $\lambda(\mathfrak{P}) \geqq \nu$, and hence $\mathfrak{P} = \mathfrak{p}^* = (x, y)$, so that if $\lambda(\mathfrak{P}) = 1$, $\mathfrak{P}$ is a v-ideal. Since $\lambda(\mathfrak{P}') < \lambda(\mathfrak{P})$ we may assume, according to our induction, that $\mathfrak{P}'$ is a v-ideal. Now $x'^\nu \mathfrak{P}'$ is the extended ideal of $\mathfrak{P}$, and from the expression of the general element t of $\mathfrak{P}$ it is seen that the subform of degree ν of t contains only the term y^ν (since $\alpha_1 > \nu$). By Theorem 4.3, our assertion that $\mathfrak{P}$ is a v-ideal, will follow, provided it is shown that $\mathfrak{P}$ *is the contracted ideal of* $x'^\nu\mathfrak{P}'$. The proof of this is immediate. The general element $\bar{t}$ of the contracted ideal of $x'^\nu\mathfrak{P}'$ is of the form $x'^\nu\bar{\tau}(x', y')$, and we can assert that not only is $\bar{\tau}$ a variable element of $\mathfrak{P}'$ (this is obvious) but also that its coefficients $\bar{\tau}_{ij}$ satisfy those inequalities (see Theorem 11.1) which insure the irreducibility of $\bar{\tau}$, *because this is true for the general element* $t = x'^\nu\tau(x', y')$ *of the ideal* $\mathfrak{P}$. It follows that $\bar{t}$ is necessarily of the form $\bar{t} = \epsilon(x, y) \prod_{i=1}^{\nu} (y - y_i)$, where $y_1 = \sum c_i x^{\alpha_i/\nu} + \sum_{j=0}^{\infty} t_j x^{(\alpha_{k+1}+j)/\nu}$, whence $\bar{t}$ is a variable element of $\mathfrak{P}$. Since $\mathfrak{P}$ is contained in the contracted ideal of $x'^\nu\mathfrak{P}'$, it follows that $\mathfrak{P}$ coincides with the contracted ideal, q. e. d.

12. The class of complete ideals. It has been pointed out in section 7 that a product of valuation ideals is not always a valuation ideal. The class of ideals (in $\mathfrak{k}[x, y]$ or in $\mathfrak{k}\{x, y\}$) which can be factored into valuation ideals is therefore larger than the class of valuation ideals. We shall call ideals in this class *complete ideals*. The term "complete" is suggested by the notion of complete linear systems in algebraic geometry, these being the linear systems which are defined uniquely by *base conditions*, i. e. by the condition of passing with assigned multiplicities through an assigned set of proper or infinitely near points. It will be seen that complete ideals are those and only those ideals whose elements are subject to given base conditions, and to no other conditions. In other words, the polynomials which belong to a complete ideal and whose degree is not greater than a given integer n, form, for any n, a complete linear system.

By our definition of complete ideals, the class of complete ideals is closed under multiplication. Moreover, by Theorem 7.1, a complete ideal has a unique factorization into simple complete ideals, these last ones being necessarily valuation ideals. Our next aim is to prove that this class is also closed under the other ideal operations $([\,,\,])$, $(:)$ (intersection and quotient); not however under addition $(+)$.* The theorem which we wish to prove is the following:

THEOREM 12.1. *If $\mathfrak{A}$ and $\mathfrak{B}$ are complete ideals and $\mathfrak{C}$ is an arbitrary ideal, then $[\mathfrak{A}, \mathfrak{B}]$ and $\mathfrak{A} : \mathfrak{C}$ are complete ideals.*

This theorem is an immediate consequence of the following:

LEMMA. *Any complete ideal is the intersection of valuation ideals, and, conversely, the intersection of valuation ideals is a complete ideal.*

Proof of the lemma. Let $\mathfrak{A} = [\mathfrak{A}_1, \mathfrak{A}_2, \cdots, \mathfrak{A}_k]$, where each $\mathfrak{A}_i$ is a valuation ideal for some valuation B_i. We may assume that the $\mathfrak{A}_i$ belong to one and the same prime ideal $\mathfrak{p} = (x, y)$. Denote by t the general element of $\mathfrak{A}$ and let $t = t_1 t_2 \cdots t_m$ be the factorization of t into irreducible factors in the ring $\mathfrak{K}\{x, y\}$, where $\mathfrak{K}$ is the algebraic closure of the field of the coefficients of t. Let $\alpha_{ij} = v(t_i)$ be the value of t_i in the valuation B_j $(i = 1, 2, \cdots, m;\ j = 1, 2, \cdots, k)$, and let $\mathfrak{A}_{ij}$ be the v-ideal belonging to B_j such that $v(\mathfrak{A}_{ij}) = \alpha_{ij}$. Since $v(t_i) = v(\mathfrak{A}_{ij})$ in B_j, t_i is a variable element of $\mathfrak{A}_{ij}$, whence t_i is also a variable element of the intersection

---

* *Example.* The ideals $(y^2, \mathfrak{p}^3)$, $(x^2, \mathfrak{p}^3)$ are v-ideals, but their sum (join) is the *simple* ideal $(x^2, y^2, \mathfrak{p}^3)$, which is obviously not a v-ideal and consequently not complete.

$$\mathfrak{B}_i = [\mathfrak{A}_{i1}, \mathfrak{A}_{i2}, \cdots, \mathfrak{A}_{ik}], \qquad (i = 1, 2, \cdots, m).$$

Let τ_i be the general element of $\mathfrak{B}_i$. We have $v(\tau_i) \leqq v(t_i)$ in B_j, since all the linear relations between the coefficients of τ_i are also satisfied by the coefficients of t_i. On the other hand, we have in any of the valuations B_j, $v(\tau_i) = v(\mathfrak{B}_i) \geqq v(\mathfrak{A}_{ij}) = v(t_i)$. Hence $v(\tau_i) = v(t_i)$. Let $\tau = \tau_1\tau_2 \cdots \tau_m$ be the direct product of the elements τ_i. The elements t and τ have the same value in each of the valuations B_j, whence $v(\tau) = v(\mathfrak{A}) \geqq v(\mathfrak{A}_j)$ in B_j. This shows that τ is a variable element of $\mathfrak{A}_j$, whence τ also is a variable element of the intersection $\mathfrak{A}$ of the $\mathfrak{A}_j$. But τ is not less general than t (since τ_i is the general element of the ideal $\mathfrak{B}_i$, of which t_i is a variable element, and since τ is a *direct* product of the τ_i); consequently, since t is the general element of $\mathfrak{A}$, also τ is the general element of $\mathfrak{A}$. We may then identify t with τ, and we deduce *that t is the direct product of its irreducible factors t_i and that t_i is the general element of $\mathfrak{B}_i$.*

We now apply the considerations developed in the proof of Theorem 11.2. For the irreducible element t_i we have the following uniformization: $t_i = \epsilon_i(x, y) \prod_{h=1}^{\nu_i} (y - y_h^{(i)})$, where

$$y_1^{(i)} = \sum_{j=1}^{k} c_{ij}x^{\alpha_{ij}/\nu_i} + \sum_{j=0}^{\infty} \tau_{ij}x^{(\alpha_{i,k+1}+j)/\nu_i},$$

τ_{i0}—transcendental with respect to field $\mathfrak{k}$.

Since the inequalities $v(t) \geqq v(\mathfrak{A}_j)$ in B_j, $j = 1, 2, \cdots, k$, are the only conditions which should be imposed on t, it follows that all the τ_{ij} are indeterminates, as the value of t_i in any algebraic valuation is not altered if we regard the τ_{ij} as indeterminates. Also $\epsilon_i(x, y)$ is to be regarded as a unit element with indeterminate coefficients, and the product of ϵ_i by $\prod_{i=1}^{\nu_i} (y - y_h^{(i)})$ is to be intended as direct. But then we must have necessarily $(\alpha_{i1}, \alpha_{i2}, \cdots, \alpha_{i,k+1}, \nu_i) = 1$ and consequently t_i is the general element of a simple v-ideal (by the preceding section, Remark 2). Hence $\mathfrak{B}_i = \mathfrak{P}^{(i)}$, where $\mathfrak{P}^{(i)}$ is a simple v-ideal. By Theorem 10.1, $t_1t_2 \cdots t_m$ is a quasi-general element of the ideal $\mathfrak{P}^{(1)}\mathfrak{P}^{(2)} \cdots \mathfrak{P}^{(m)}$. But the coefficients of the direct product Πt_i, i. e. of t, satisfy only linear relations, since t is the general element of $\mathfrak{A}$. Hence t is also the general element of $\mathfrak{P}^{(1)}\mathfrak{P}^{(2)} \cdots \mathfrak{P}^{(m)}$, and consequently $\mathfrak{A} = \mathfrak{P}^{(1)}\mathfrak{P}^{(2)} \cdots \mathfrak{P}^{(m)}$. This proves the first part of the lemma.

To prove the second part of the lemma, let $\mathfrak{A} = \mathfrak{A}_1\mathfrak{A}_2 \cdots \mathfrak{A}_m$ be a complete ideal, where $\mathfrak{A}_1, \mathfrak{A}_2, \cdots, \mathfrak{A}_m$ are simple v-ideals, which we may assume as primary ideals belonging to one and the same prime ideal $\mathfrak{p} = (x, y)$. Let

$\mathfrak{A}_j$ be of kind h_j, and let $h = \max(h_1, h_2, \cdots, h_m)$. The assertion of the lemma is trivial if $m = 1$. We assume that the assertion is true for all integers h' such that $h' < h$ (if $h = 1$, then necessarily $m = 1$). Let $\mathfrak{A}_j$ belong to a valuation B_j. Assuming that $v(y) \geqq v(x)$ in each of the valuation B_j, we apply the quadratic transformation T, $x' = x$, $y' = y/x$ to the polynomial ring $\mathfrak{O} = \mathfrak{k}[x, y]$, getting the ring $\mathfrak{O} = \mathfrak{k}[x', y']$ of polynomials in x', y'. Let $\mathfrak{O}'\mathfrak{A}_j = x^{\rho_j}\mathfrak{A}'_j$, whence $\mathfrak{O}'\mathfrak{A} = x'^{\rho}\mathfrak{A}'_1\mathfrak{A}'_2 \cdots \mathfrak{A}'_m$, $\rho = \sum_{j=1}^{m} \rho_j$, where $\mathfrak{A}_j \equiv 0(\mathfrak{p}^{\rho_j})$ and $\mathfrak{A}_j \not\equiv 0(\mathfrak{p}^{\rho_j+1})$. By Theorem 4. 1, $\mathfrak{A}'_j$ is a v-ideal in $\mathfrak{O}'$, and by Theorems 5. 2 and 5. 3 $\mathfrak{A}'_j$ is a simple v-ideal of kind $h_j - 1$. The product $\prod_{i-1}^{\rho} \mathfrak{A}'_j$ can be first written as the intersection of the partial products consisting of factors $\mathfrak{A}'_j$ which belong to one and the same prime ideal. Then, by our induction, each partial product can be written as the intersection of v-ideals Let then

$$\mathfrak{A}' = \mathfrak{A}'_1\mathfrak{A}'_2 \cdots \mathfrak{A}'_m = [\mathfrak{B}'_1, \mathfrak{B}'_2, \cdots, \mathfrak{B}'_k],$$

whence

$$\mathfrak{O}'\mathfrak{A} = [x^{\rho}\mathfrak{B}'_1, x^{\rho}\mathfrak{B}'_2, \cdots, x^{\rho}\mathfrak{B}'_k], \tag{28}$$

where $\mathfrak{B}'_1, \mathfrak{B}'_2, \cdots, \mathfrak{B}'_k$ are v-ideals in $\mathfrak{O}'$.

Let σ_i be the smallest integer such that $x'^{\sigma_i}\mathfrak{B}'_i$ is an extended ideal of an ideal in $\mathfrak{O}$, and let $\mathfrak{B}_i$ be the contracted ideal of $x'^{\sigma_i}\mathfrak{B}'_i$. By Theorem 4. 3, $\mathfrak{B}_i$ is a v-ideal and $\mathfrak{B}_i = T^{-1}(\mathfrak{B}'_i)$. We have $\mathfrak{O}'\mathfrak{B}_i = x'^{\sigma_i}\mathfrak{B}'_i$ and $\mathfrak{O}'\mathfrak{A} = x^{\rho}\mathfrak{A}'$. We assert that $\sigma_i \leqq \rho$. In fact, assuming that $\sigma_i > \rho$, we have

$$\mathfrak{O}'\mathfrak{p}^{\sigma_i-\rho}\mathfrak{A} = x'^{\sigma_i}\mathfrak{A}' \subseteqq x'^{\sigma_i}\mathfrak{B}'_i,$$

whence, passing to the contracted ideal of $x'^{\sigma_i}\mathfrak{B}'_i$, $\mathfrak{p}^{\sigma_i-\rho}\mathfrak{A} \equiv 0(\mathfrak{B}_i)$. Now $\mathfrak{A} \equiv 0(\mathfrak{p}^{\rho})$ and $\mathfrak{A} \not\equiv 0(\mathfrak{p}^{\rho+1})$, hence $\mathfrak{p}^{\sigma_i-\rho}\mathfrak{A} \equiv 0(\mathfrak{p}^{\sigma_i})$ and $\not\equiv 0(\mathfrak{p}^{\sigma_i+1})$. Since also $\mathfrak{B}_i$ is in $\mathfrak{p}^{\sigma_i}$ and not in $\mathfrak{p}^{\sigma_i+1}$, the congruence $\mathfrak{p}^{\sigma_i-\rho}\mathfrak{A} \equiv 0(\mathfrak{B}_i)$ implies that the subforms of degree σ_i of the polynomials in $\mathfrak{B}_i$ form a linear system $\Omega(\mathfrak{B}_i)$ of dimension $\geqq \sigma_i - \rho > 0$. This is impossible since, by our definition of $\mathfrak{B}_i (= T^{-1}(\mathfrak{B}'_i))$, the system $\Omega(\mathfrak{B}_i)$ is of dimension zero.

We have therefore $\sigma_i \leqq \rho$, and consequently $x^{\rho}\mathfrak{B}'_i$ is an extended ideal, its contracted ideal being $\mathfrak{p}^{\rho-\sigma_i}\mathfrak{B}_i$ (Theorem 4. 3). Denoting by $\mathfrak{A}^*$ the contracted ideal of $\mathfrak{O}'\mathfrak{A}$, we obtain from (28):

$$\mathfrak{A} \subseteqq \mathfrak{A}^* = [\mathfrak{p}^{\rho-\sigma_1}\mathfrak{B}_1, \mathfrak{p}^{\rho-\sigma_2}\mathfrak{B}_2, \cdots, \mathfrak{p}^{\rho-\sigma_k}\mathfrak{B}_k].$$

Now $\mathfrak{B}_i$ is a valuation ideal for some valuation, and in that valuation the ideal $\mathfrak{p}^{\rho-\sigma_i}\mathfrak{B}_i$ is equivalent to some valuation ideal $\mathfrak{C}_i$. By Corollary 3.3 we have $\mathfrak{p}^{\rho-\sigma_i}\mathfrak{B}_i = [\mathfrak{C}_i, \mathfrak{p}^{\rho}]$, whence

$$\mathfrak{A} \subseteqq \mathfrak{A}^* = [\mathfrak{C}_1, \mathfrak{C}_2, \cdots, \mathfrak{C}_k, \mathfrak{p}^{\rho}].$$

Hence $\mathfrak{A}^*$ is the intersection of valuation ideals.[13] By the first part of the lemma, just proved, $\mathfrak{A}^*$ is a complete ideal,

$$\mathfrak{A}^* = \mathfrak{A}^*_1\mathfrak{A}^*_2 \cdots \mathfrak{A}^*_n,$$

where the $\mathfrak{A}^*_j$ are simple v-ideals. $\mathfrak{A}$ and $\mathfrak{A}^*$ have the same extended ideal $x'^{\rho}\mathfrak{A}'$. If then $T(\mathfrak{A}^*_j) = \mathfrak{A}'^*_j$ we must have

$$x'^{\rho}\mathfrak{A}'_1\mathfrak{A}'_2 \cdots \mathfrak{A}'_m = \mathfrak{O}'\mathfrak{A} = \mathfrak{O}'\mathfrak{A}^* = x'^{\rho}\mathfrak{A}'^*_1\mathfrak{A}'^*_2 \cdots \mathfrak{A}'^*_n.$$

By the unique factorization theorem (7.1), it follows, for a proper ordering of the factors, $m = n$, $\mathfrak{A}'_j = \mathfrak{A}'^*_j$, whence also $\mathfrak{A}_j = \mathfrak{A}^*_j$, $\mathfrak{A} = \mathfrak{A}^*$, and consequently $\mathfrak{A}$ is the intersection of v-ideals, q. e. d.

Theorem 12.1 now follows immediately. That $[\mathfrak{A}, \mathfrak{B}]$ is a complete ideal, if $\mathfrak{A}$ and $\mathfrak{B}$ are complete ideals, is now trivial, since, by our lemma, complete ideals can also be defined as intersections of valuation ideals. Consider now the ideal $\mathfrak{A} : \mathfrak{C}$, where $\mathfrak{A}$ is a complete ideal and $\mathfrak{C}$ is an arbitrary ideal. We have $\mathfrak{A} = [\mathfrak{A}_1, \mathfrak{A}_2, \cdots, \mathfrak{A}_k]$, where the $\mathfrak{A}_j$ are valuation ideals, whence $\mathfrak{A} : \mathfrak{C} = [\mathfrak{A}_1 : \mathfrak{C}, \mathfrak{A}_2 : \mathfrak{C}, \cdots, \mathfrak{A}_k : \mathfrak{C}]$. But by Theorem 2.1, the quotient $\mathfrak{A}_j : \mathfrak{C}$ is again a valuation ideal. Hence $\mathfrak{A} : \mathfrak{C}$ is the intersection of valuation ideals, and therefore is a complete ideal, q. e. d.

COROLLARY 12.2. *If $\mathfrak{A}$ is a complete ideal* (in $\mathfrak{k}[x, y]$ or in $\mathfrak{k}\{x, y\}$) *and if $x'^{\rho}\mathfrak{A}'$ is the extended ideal $\mathfrak{O}'\mathfrak{A}$ in the ring $\mathfrak{k}[x', y']$ or $\mathfrak{k}\{x', y'\}$, where $x' = x$, $y' = y/x$, then $\mathfrak{A}'$ is a complete ideal and $\mathfrak{A}$ is the contracted ideal of $x'^{\rho}\mathfrak{A}'$.*

That $\mathfrak{A}'$ is a complete ideal is trivial. The second part of the corollary follows from the relation $\mathfrak{A} = \mathfrak{A}^*$ established in the course of the proof of the second part of the lemma.

---

[13] *Any power of $\mathfrak{p}$ is a valuation ideal.* Consider the 0-dimensional valuation defined by the divisor (x') and by the point $y' = 0$ of this divisor, i. e. if $F(x', y')$ is any rational function of x', y', then put $v(F) = (m, n)$, where x'^m is the highest power of x' which divides F, $F = x'^m F_1(x'y')$, $F_1(0, y') \neq 0$, and where y'^n is the highest power of y' which divides $F_1(0, y')$. Then it is easily seen that the sequence of v-ideals

When a linear system of curves $f=0$ is subjected to base conditions, it is required that the curves of the system have assigned intersection multiplicities with an assigned set of algebraic branches γ_j. For each branch γ_j the corresponding condition is equivalent to the condition that the value of the polynomial f in the valuation B_j defined by the branch γ_j be not inferior to a given integer; in other words: it is required that $f \equiv 0(\mathfrak{A}_j)$, where $\mathfrak{A}_j$ is a given valuation ideal belonging to B_j. The full set of base conditions is then described by the congruence: $f \equiv 0([\mathfrak{A}_1, \mathfrak{A}_2, \cdots])$, i. e. by the condition that f belong to a given complete ideal. However, the representation of a complete ideal as an intersection of valuation ideals is not unique. For instance, let $\mathfrak{A} = (xy, \mathfrak{p}^3)$, where $\mathfrak{p} = (x, y)$. $\mathfrak{A}$ is not a valuation ideal, since it has only one subform xy and this is not a power of a linear form. Let $\mathfrak{A}_1 = (x^2, xy, \mathfrak{p}^3)$, $\mathfrak{A}_2 = (xy, y^2, \mathfrak{p}^3)$, $\mathfrak{A}'_1 = (x, x^2, xy, \mathfrak{p}^3)$, $\mathfrak{A}'_2 = (y, xy, y^2, \mathfrak{p}^3)$. These 4 ideals are valuation ideals and we have $\mathfrak{A} = [\mathfrak{A}_1, \mathfrak{A}_2] = [\mathfrak{A}'_1, \mathfrak{A}'_2]$. This ambiguity is the algebraic equivalent of the distinction which the geometric theory makes between the *assigned* or *virtual* multiplicities of the curves of a linear system and the *effective* multiplicities of these curves. It is the representation of a complete ideal as a *product* of valuation ideals that is unique and puts into evidence the effective multiplicities of the general curve of a linear system.*

We point out explicitly the following result arrived at in the course of the first part of the above proof: *the general element t of a complete ideal $\mathfrak{A}$ is the direct product of its irreducible factors, and the factorization of t yields the factorization of $\mathfrak{A}$ into simple v-ideals.* This is the generalization to complete ideals of the similar property of valuation ideals given in Theorem 11. 2.

We conclude this section with the definition of an operation which assigns to each ideal $\mathfrak{A}$ in $\mathfrak{O}$ a uniquely determined complete ideal $\mathfrak{A}'$, *the complete ideal determined by* $\mathfrak{A}$. The analogue of this operation in the theory of linear systems is given by the passage from an arbitrary linear system to the corresponding complete linear system. *We define $\mathfrak{A}'$ as the intersection of all the complete ideals containing* $\mathfrak{A}$. Since $\mathfrak{A}$ has a finite length, it is clear that $\mathfrak{A}'$ is also the intersection of a finite number of complete ideals, *hence $\mathfrak{A}'$ itself is a complete ideal,* i. e. *$\mathfrak{A}'$ is the smallest complete ideal containing* $\mathfrak{A}$. The operation comes under the heading of the (′) *operations* studied in other connections by van der Waerden ([9], § 103), Prüfer[8] and others, since it enjoys the following formal properties:

---

in $\mathfrak{k}[x, y]$ consists of all ideals of the form $(x^\lambda y^{\rho-\lambda}, x^{\lambda-1} y^{\rho-\lambda+1}, \ldots, y^\rho, \mathfrak{p}^{\rho+1})$, $\rho \geqq \lambda \geqq 0$, $\rho = 1, 2, 3, \ldots$. For $\lambda = \rho$, we find the ideal $\mathfrak{p}^\rho$.

* See Note at end of Section 12.

1. $(\mathfrak{A}')' = \mathfrak{A}'$. 2. If $\mathfrak{A}_1 \supseteqq \mathfrak{A}_2$, then $\mathfrak{A}'_1 \supseteqq \mathfrak{A}'_2$.
3. $(\mathfrak{A}'_1, \mathfrak{A}'_2)' = (\mathfrak{A}_1, \mathfrak{A}_2)'$. 4. $(\mathfrak{A}'_1\mathfrak{A}'_2)' = (\mathfrak{A}_1\mathfrak{A}_2)'$.
5. $(a)' = (a)$; $((a)\cdot\mathfrak{A})' = (a)\cdot\mathfrak{A}'$.

Proof. 1. Trivial, because $\mathfrak{A}'$ is a complete ideal itself;

2. Self-evident;

3. By 2, $(\mathfrak{A}_1, \mathfrak{A}_2)' \supseteqq (\mathfrak{A}'_1, \mathfrak{A}'_2)$, whence $(\mathfrak{A}_1, \mathfrak{A}_2)' \supseteqq (\mathfrak{A}'_1, \mathfrak{A}'_2)'$. On the other hand $\mathfrak{A}'_1 \supseteqq \mathfrak{A}_1$, $\mathfrak{A}'_2 \supseteqq \mathfrak{A}_2$, whence $(\mathfrak{A}'_1, \mathfrak{A}'_2)' \supseteqq (\mathfrak{A}_1, \mathfrak{A}_2)'$, by 2.

4. By 2, we have $(\mathfrak{A}'_1\mathfrak{A}'_2)' \supseteqq (\mathfrak{A}_1\mathfrak{A}_2)'$. Let $\mathfrak{B}$ be any valuation ideal belonging to some valuation B and containing $\mathfrak{A}_1\mathfrak{A}_2$, and let $\mathfrak{B}_i$, $i = 1, 2$, be the valuation ideal for B which is equivalent to $\mathfrak{A}_i$. We have $v(\mathfrak{B}_1\mathfrak{B}_2) = v(\mathfrak{A}_1\mathfrak{A}_2) \geqq v(\mathfrak{B})$, whence $\mathfrak{B}_1\mathfrak{B}_2 \equiv 0(\mathfrak{B})$. Since $\mathfrak{A}_i \equiv 0(\mathfrak{B}_i)$ and $\mathfrak{B}_i$ is a valuation ideal, hence complete, we have $\mathfrak{A}'_i \equiv 0(\mathfrak{B}_i)$, consequently $\mathfrak{A}'_1\mathfrak{A}'_2 \equiv 0(\mathfrak{B}_1\mathfrak{B}_2)$. Since every complete ideal is the intersection of valuation ideals, it follows that $(\mathfrak{A}_1\mathfrak{A}_2)'$ is the intersection of all the valuation ideals $\mathfrak{B}$ containing $\mathfrak{A}_1\mathfrak{A}_2$. Hence $\mathfrak{A}'_1\mathfrak{A}'_2 \subseteqq (\mathfrak{A}_1\mathfrak{A}_2)'$. As a consequence also $(\mathfrak{A}'_1\mathfrak{A}'_2)' \subseteqq ((\mathfrak{A}_1\mathfrak{A}_2)')$, whence 4 follows.

We observe, however, that $(\mathfrak{A}'_1\mathfrak{A}'_2)' = \mathfrak{A}'_1\mathfrak{A}'_2$, since the product of the complete ideals $\mathfrak{A}'_1, \mathfrak{A}'_2$ is itself complete. Hence 4 can be written as follows:

4′. $\mathfrak{A}'_1\mathfrak{A}'_2 = (\mathfrak{A}_1\mathfrak{A}_2)'$.

5. A principal ideal (a) is itself a complete ideal. If $a = a_1^{\rho_1}a_2^{\rho_2}\cdots a_k^{\rho_k}$, where $a_i(x, y)$ is an irreducible polynomial, then $(a) = (a_1^{\rho_1}) \cap (a_2^{\rho_2}) \cap \cdots \cap (a_k^{\rho_k})$, and evidently $(a_i^{\rho_i})$ is a v-ideal, belonging to the 1-dimensional valuation defined by the divisor (a_i). Hence $(a)' = (a)$. In a similar straightforward manner the relation $((a)\cdot\mathfrak{A})' = (a)\mathfrak{A}'$ can be proved.

THEOREM 12. 3 (*invariance of the operation* (′) *under quadratic transformations*). *If* $\mathfrak{p} = (x, y)$ *and if* $\mathfrak{A} \equiv 0(\mathfrak{p}^{\rho})$, $\mathfrak{A} \not\equiv 0(\mathfrak{p}^{\rho+1})$, *then also* $\mathfrak{A}' \equiv 0(\mathfrak{p}^{\rho})$, $\mathfrak{A}' \not\equiv 0(\mathfrak{p}^{\rho+1})$. *Moreover, if* $\mathfrak{B}$ *is the transform of* $\mathfrak{A}$ *under the quadratic transformation* T, *then* $\mathfrak{B}'$ *is the transform of* $\mathfrak{A}'$ *under* T, i. e. $\mathfrak{O}'\mathfrak{A} = x'^{\rho}\mathfrak{B}$ implies $\mathfrak{O}'\mathfrak{A}' = x'^{\rho}\mathfrak{B}'$.

Proof. Since any power of $\mathfrak{p}$ is a complete ideal (even a valuation ideal) and since $\mathfrak{A}'$ is the smallest complete ideal containing $\mathfrak{A}$, the congruence $\mathfrak{A} \equiv 0(\mathfrak{p}^{\rho})$ implies the congruence $\mathfrak{A}' \equiv 0(\mathfrak{p}^{\rho})$. Moreover, if $\mathfrak{A} \not\equiv 0(\mathfrak{p}^{\rho+1})$, also $\mathfrak{A}' \not\equiv 0(\mathfrak{p}^{\rho+1})$ since $\mathfrak{A} \equiv 0(\mathfrak{A}')$. Let then $\mathfrak{O}'\mathfrak{A} = x'^{\rho}\mathfrak{B}$, $\mathfrak{O}'\mathfrak{A}' = x'^{\rho}\mathfrak{B}^*$, where $\mathfrak{O}' = \mathfrak{k}[x', y']$ (or $\mathfrak{O}' = \mathfrak{k}\{x', y'\}$), $x' = x$, $y' = y/x$, and where $\mathfrak{B}$ and $\mathfrak{B}^*$ are not divisible by x'. Since $\mathfrak{A} \equiv 0(\mathfrak{A}')$, also $\mathfrak{B} \equiv 0(\mathfrak{B}^*)$, and by Corollary 12. 2, $\mathfrak{B}^*$ is a complete ideal. To prove that $\mathfrak{B}^* = \mathfrak{B}'$, we have to show that if $\mathfrak{B}_1$ is any complete ideal containing $\mathfrak{B}$, then $\mathfrak{B}^* \equiv 0(\mathfrak{B}_1)$. Let σ be the smallest integer such that $x'^{\sigma}\mathfrak{B}_1$ is an extended ideal of an ideal in $\mathfrak{O}$. The same reasoning employed in the proof of the second part of the Lemma, for the derivation of the inequality $\sigma_i \leqq \rho$, can be used also now in order to show that $\sigma \leqq \rho$. It is only necessary to observe that by Corollary 12. 2 the contracted ideal of $x'^{\sigma}\mathfrak{B}_1$ is a complete ideal, say $\mathfrak{C}$. The system $\Omega(\mathfrak{C})$ of the subforms (of degree σ) of $\mathfrak{C}$ can be of dimension greater than zero, only if in the factorization of $\mathfrak{C}$ into simple v-ideals there occurs the factor $\mathfrak{p}$. We would have then $\mathfrak{C} = \mathfrak{p}\mathfrak{D}$, whence $\mathfrak{O}'\mathfrak{D} = x'^{\sigma-1}\mathfrak{B}_1$, in contradiction with the definition of the integer σ. Hence $\Omega(\mathfrak{C})$ is of dimension zero, and it was this value of the dimension of $\Omega(\mathfrak{B}_i)$ that played a rôle in the proof of the Lemma.

Since $\sigma \leqq \rho$ and $\mathfrak{B}_1 \supseteqq \mathfrak{B}$, we have $x'^{\rho}\mathfrak{B}_1 \supseteqq x'^{\rho}\mathfrak{B}$. Now let $\mathfrak{C}$ be the contracted ideal of $x'^{\sigma}\mathfrak{B}_1$. $\mathfrak{C}$ is a complete ideal, and its extension ideal is $x'^{\sigma}\mathfrak{B}_1$. Hence the extended ideal of $\mathfrak{p}^{\rho-\sigma}\mathfrak{C}$ is $x'^{\rho}\mathfrak{B}_1$, and consequently (Corollary 12. 2) $\mathfrak{p}^{\rho-\sigma}\mathfrak{C}$ is the contracted ideal of $x'^{\rho}\mathfrak{B}_1$, since $\mathfrak{p}^{\rho-\sigma}$ is also a complete ideal. Since the contracted ideal of $x'^{\rho}\mathfrak{B}$ contains $\mathfrak{A}$, the congruence $x'^{\rho}\mathfrak{B}_1 \supseteqq x'^{\rho}\mathfrak{B}$ implies the congruence $\mathfrak{p}^{\rho-\sigma}\mathfrak{C} \supseteqq \mathfrak{A}$. Hence $\mathfrak{p}^{\rho-\sigma}\mathfrak{C} \supseteqq \mathfrak{A}'$, and passing to the extended ideals in $\mathfrak{O}'$, we find $x'^{\rho}\mathfrak{B}_1 \supseteqq x'^{\rho}\mathfrak{B}^*$, i. e. $\mathfrak{B}_1 \supseteqq \mathfrak{B}^*$, q. e. d.

NOTE.—For those not familiar with the geometric terminology we give here the definition of the effective multiplicities on the basis of the present treatment. We associate with each simple v-ideal $\mathcal{P}_{k+1}$ of kind $k+1$, belonging to the prime ideal $\mathcal{P}_1 = \mathfrak{p} = (x, y)$, a point 0_{k+1} in the k-th neighborhood of the point $0_1(0, 0)$ of the (x, y)-plane. Let $\mathcal{P}_1, \mathcal{P}_2, \cdots, \mathcal{P}_k$ be the simple v-ideals of kind $1, 2, \cdots, k$ determined by $\mathcal{P}_{k+1}$ and preceding it (Theorem 6. 1), and let $0_1, 0_2, \cdots, 0_k$ be the associated points. We proceed to define the *set of base points* of the ideal $\mathcal{P}_{k+1}$ or the symbol

$$B(\mathcal{P}_{k+1}) = (0_1^{r_1} 0_2^{r_2} \cdots 0_k^{r_k} 0_{k+1}), \qquad r_i > 0,$$

and we shall say that r_i is the *effective multiplicity* of the ideal $\mathcal{P}_{k+1}$ at $0_i (r_{k+1} = 1)$. We set $B(\mathcal{P}_1) = (0_1)$. For any k we define $B(\mathcal{P}_{k+1})$ by induction with respect to k. We know, by Theorem 6. 2, that if B is any valuation for which $\mathcal{P}_{k+1}$ is a v-ideal, the v-ideals for B which precede $\mathcal{P}_{k+1}$ are independent of B. Let $\mathfrak{q}_\lambda$ be the v-ideal which is followed immediately by $\mathcal{P}_{k+1}$ and let $\mathfrak{q}_\lambda = \mathcal{P}_1^{a_1} \mathcal{P}_2^{a_2} \cdots \mathcal{P}_k^{a_k}$. Assuming that the symbol $B(\mathcal{P}_i)$ has already been defined for all $i < k+1$, we put

$$B(\mathcal{P}_{k+1}) = [B(\mathcal{P}_1)]^{a_1} [B(\mathcal{P}_2)]^{a_2} \cdots [B(\mathcal{P}_k)]^{a_k} O_{k+1},$$

where, if

$$B(\mathcal{P}_{i+1}) = (0_1^{s_1} 0_2^{s_2} \cdots 0_i^{s_i} 0_{i+1})$$

and

$$B(\mathfrak{P}_{j+1}) = (0_1^{t_1} 0_2^{t_2} \cdots 0_j^{t_j} 0_{j+1}),$$

then

$$B(\mathfrak{P}_i) B(\mathfrak{P}_j) = (0_1^{s_1+t_1} 0_2^{s_2+t_2} \cdots).$$

To check this definition against the customary geometric definition, we point out two implications of our definition:

(1) $\mathfrak{P}_{k+1}$ *is exactly divisible by* $\mathfrak{p}^{r_1}$, i. e. $\mathfrak{P}_{k+1} \equiv 0(\mathfrak{p}^{r})$, $\mathfrak{P}_{k+1} \not\equiv 0(\mathfrak{p}^{r+1})$. The assertion is true for $k = 0$. Assuming that is true for $\mathfrak{P}_{i+1}$, $i = 0, 1, \cdots, k-1$ and that $\mathfrak{P}_{i+1}$ is exactly divisible by $\mathfrak{p}^{r_{i1}}$, then

$$B(\mathfrak{P}_{i+1}) = (0_1^{r_{i1}} \cdots), \qquad r_1 = \alpha_1 r_{01} + \cdots + \alpha_k r_{k-1,1}$$

and from this we conclude, in view of $\mathfrak{q}_\lambda = \mathfrak{P}_1^{\alpha_1} \mathfrak{P}_2^{\alpha_2} \cdots \mathfrak{P}_k^{\alpha_k}$, that $\mathfrak{q}_\lambda$ is exactly divisible by $\mathfrak{p}^{r_1}$. Now $\mathfrak{p}\mathfrak{q}_\lambda \equiv 0(\mathfrak{P}_{k+1})$, since $\mathfrak{P}_{k+1}$ is a maximal subideal of $\mathfrak{q}_\lambda$. Hence $\mathfrak{P}_{k+1}$ is divisible at most by $\mathfrak{p}^{r_1+1}$. If $\mathfrak{P}_{k+1}$ was divisible by $\mathfrak{p}^{r_1+1}$, then the system of subforms $\Omega(\mathfrak{p}\mathfrak{q}_r)$ would have been a subsystem of $\Omega(\mathfrak{P}_{k+1})$. This is impossible, since $\Omega(\mathfrak{p}\mathfrak{q}_\lambda)$ is of dimension $\geqq 1$, while $\Omega(\mathfrak{P}_{k+1})$ is of dimension 0. Hence $\mathfrak{P}_{k+1}$ is divisible exactly by $\mathfrak{p}^{r_1}$. We have thus proved that in the general polynomial $f(x, y)$ in $\mathfrak{P}_{k+1}$ the terms of lowest degree are of degree r_1, i. e. *the curve* $f = 0$ *has at* 0_r *an* r_1*-fold point* (*while no curve* $f = 0$, $f \varepsilon \mathfrak{P}_{k+1}$, *has at* 0_1 *a multiplicity less than* r_1).

(2) *If* $\mathfrak{P}'_i$ *is the transform of* $\mathfrak{P}_{i+1}$ *by a quadratic transformation* T *having at* 0_1 *a fundamental point, then* $B(\mathfrak{P}'_k) = (0'^{r_2}_1 \cdots 0'^{r_k}_{k-1} 0'_k)$, *where* $0'_i$ *is the point associated with* $\mathfrak{P}'_i$. To prove this, we first observe that from the fact, just proved above, that $\mathfrak{q}_\lambda$ and $\mathfrak{P}_{k+1}$ are both divisible exactly by the same power, $\mathfrak{p}^{r_1}$, of $\mathfrak{p}$, it follows that $\Omega(\mathfrak{P}_{k+1})$ is a subsystem of $\Omega(\mathfrak{q}_\lambda)$. Since $\mathfrak{P}_{k+1}$ is a maximal subideal of $\mathfrak{q}_\lambda$, the dimension of $\Omega(\mathfrak{q}_\lambda)$ cannot exceed the dimension of $\Omega(\mathfrak{P}_{k+1})$ augmented by 1. But $\Omega(\mathfrak{P}_{k+1})$ is of dimension 0, while $\Omega(\mathfrak{q}_\lambda)$ is of dimension α_1. Hence $\alpha_1 \leqq 1$. Let $T(\mathfrak{q}_\lambda) = x'^{r_1} \mathfrak{q}'_\mu$, where $\mathfrak{q}'_\mu = \mathfrak{P}'^{\alpha_2}_1 \mathfrak{P}'^{\alpha_3}_2 \cdots \mathfrak{P}'^{\alpha_k}_{k-1}$. We assert that $\mathfrak{q}'_\mu$ *is the immediate predecessor of* $\mathfrak{P}'_k$. If $\alpha_1 = 0$, the assertion follows immediately from Theorem 4. 4, since we have in this case $T^{-1}(\mathfrak{q}'_\mu) = \mathfrak{q}_\lambda$. Let now $\alpha_1 = 1$. Assume that $\mathfrak{q}'_\mu$ is not the immediate predecessor of $\mathfrak{P}'_k$ and let $\mathfrak{q}'_s$ be a v-ideal between $\mathfrak{q}'_\mu$ and $\mathfrak{P}'_k$. Let $\mathfrak{q}_m = T^{-1}(\mathfrak{q}'_\mu) = \mathfrak{P}_2^{\alpha_2} \cdots \mathfrak{P}_k^{\alpha_k}$, $\mathfrak{q}_r = T^{-1}(\mathfrak{q}'_s)$, whence $\mathfrak{q}_m \supset \mathfrak{q}_r \supset \mathfrak{P}_{k+1}$ (by Theorem 4. 4). We have $\mathfrak{q}_r \equiv 0(\mathfrak{q}_m) \equiv 0(\mathfrak{p}^{r_1-1})$, and also $\mathfrak{q}_r \not\equiv 0(\mathfrak{p}^{r_1+1})$, since $\mathfrak{P}_{k+1} \equiv 0(\mathfrak{q}_r)$. We also have $\mathfrak{p}\mathfrak{q}_m = \mathfrak{q}_\lambda \equiv 0(\mathfrak{q}_r)$, since $\mathfrak{q}_\lambda$ is the immediate predecessor of $\mathfrak{P}_{k+1}$. If $\mathfrak{q}_r$ was divisible by $\mathfrak{p}^{r_1}$, then $\Omega(\mathfrak{q}_\lambda)$ would be a subsystem of $\Omega(\mathfrak{q}_r)$, and this is impossible, since $\Omega(\mathfrak{q}_\lambda)$ is of dimension 1, while $\Omega(\mathfrak{q}_r)$ must be of dimension 0. Hence $\mathfrak{q}_r$ is divisible exactly by $\mathfrak{p}^{r_1-1}$. Now we have $v(\mathfrak{q}_r) > v(\mathfrak{q}_m)$, $v(\mathfrak{p}\mathfrak{q}_r) > v(\mathfrak{p}\mathfrak{q}_m) = v(\mathfrak{q}_\lambda)$, and consequently $\mathfrak{p}\mathfrak{q}_r \equiv 0(\mathfrak{P}_{k+1})$. This is a contradiction, since both $\mathfrak{p}\mathfrak{q}_r$ and $\mathfrak{P}_{k+1}$ are divisible exactly by $\mathfrak{p}^{r_1}$ and $\Omega(\mathfrak{p}\mathfrak{q}_r)$ is of dimension 1. It is thus proved that $\mathfrak{q}'_\mu$ is the immediate predecessor of $\mathfrak{P}'_k$. As a consequence we have

$$B(\mathfrak{P}'_k) = [B(\mathfrak{P}'_1)]^{\alpha_2} [B(\mathfrak{P}'_2)]^{\alpha_3} \cdots [B(\mathfrak{P}'_{k-1})]^{\alpha_k} 0'_k.$$

From this relation our statement follows immediately by induction with respect to k. Our result implies that *if the general curve* $f = 0$, $f \varepsilon \mathfrak{P}_{k+1}$, *passes through* $0_1, 0_2, \cdots, 0_{k+1}$ *with effective multiplicities* $r_1, r_2, \cdots, r_k, 1$, *then its transform by* *the quadratic transformation T passes through the points* $0'_1, 0'_2, \cdots, 0'_k$ *with effective multiplicities* $r_2, r_3, \cdots, r_k, 1$. The identity between our definition of effective multiplicities and the customary geometric definition is thus fully proved. It is hardly necessary to add that the definition of the symbol $B(\mathfrak{A})$ for any complete ideal amounts to postulating the relation $B(\mathfrak{B}\mathfrak{C}) = B(\mathfrak{B})B(\mathfrak{C})$.

13. Simple v-ideals and divisors of the second kind. Let $\mathfrak{P}$ be a divisor of the field $\Sigma = \mathfrak{k}(x, y)$, i. e. an homomorphism of Σ upon a field Σ' of dimension 1 (and the symbol ∞), and let us assume that $\mathfrak{P}$ is of second kind for the ring $\mathfrak{O} = \mathfrak{k}[x, y]$, i. e. that the prime ideal $\mathfrak{p}$ determined in $\mathfrak{O}$ by the divisor $\mathfrak{P}$ is zero-dimensional, say $\mathfrak{p} = (x, y)$ (we exclude the case in which x or y are mapped upon ∞). The points of the Riemann surface of the field Σ' define a set of valuations $\{B_\alpha\}$ of the field Σ, all of rank 2. Let $\{\mathfrak{q}_i^{(\alpha)}\}$ be the Jordan sequence of v-ideals in $\mathfrak{O}$ belonging to the valuation B_α. The ideal $\mathfrak{q}_1^{(\alpha)} = \mathfrak{p}$ is independent of α. There may be other values of i such that $\mathfrak{q}_i^{(\alpha)}$ is independent of α, and, in particular, there may occur in the sequence $\{\mathfrak{q}_i^{(\alpha)}\}$ simple ideals independent of α. Their number is necessarily finite, because the simple v-ideals in the sequence $\{\mathfrak{q}_i^{(\alpha)}\}$ determine the sequence completely (Theorem 6. 2) and since the valuations B_α are all distinct. Let $\mathcal{P}_\rho$ be the last simple v-ideal, of kind ρ, which occurs in all the sequences $\{\mathfrak{q}_i^{(\alpha)}\}$. The simple ideal $\mathcal{P}^{(\alpha)}_{\rho+1}$, of kind $\rho + 1$, will then vary with α. We have thus associated with every divisor of Σ, of the second kind, with respect to $\mathfrak{O}$, a simple v-ideal $\mathcal{P}_\rho$ in $\mathfrak{O}$. If we apply ρ successive quadratic transformations, getting a polynomial ring $\mathfrak{k}[X, Y]$ of the new indeterminates X, Y, the ideal $\mathcal{P}^{(\alpha)}_{\rho+1}$ is transformed into a prime 0-dimensional ideal $\mathfrak{p}^{(\alpha)} = (X, Y - c^{(\alpha)})$, and the constant $c^{(\alpha)}$ must vary as α varies. As a consequence, the divisor $\mathfrak{P}$ is of the first kind with respect to the ring $\mathfrak{k}[X, Y]$, and the corresponding prime ideal in this ring is necessarily the 1-dimensional ideal (X). This shows that there exists a divisor $\mathfrak{P}$ for any preassigned simple v-ideal $\mathcal{P}_\rho$ and that $\mathfrak{P}$ is uniquely determined by $\mathcal{P}_\rho$. *We have then a one to one correspondence between the divisors of the second kind (with respect to the ring* $\mathfrak{k}[x, y]$*) and the* 0*-dimensional simple v-ideals in* $\mathfrak{k}[x, y]$. The field Σ' upon which Σ is mapped by $\mathfrak{P}$ coincides with the field $\mathfrak{k}(Y)$ and is therefore purely transcendental.

It is important to point out that as the valuation B_α runs through the set $\{B_\alpha\}$ determined by the points of the field Σ', the set $\{\mathcal{P}^{(\alpha)}_{\rho+1}\}$ will include *all* the simple v-ideals $\mathcal{P}_{\rho+1}$ of kind $\rho + 1$, such that $\mathcal{P}_{\rho+1}$ and $\mathcal{P}_\rho$ belong together to one and the same valuation. This follows from the fact, that all such ideals $\mathcal{P}_{\rho+1}$ are transformed by ρ successive quadratic transformations into the above ideals $\mathfrak{p}^{(\alpha)}$.

The preceding considerations refer to the field $\mathfrak{k}(x, y)$ of rational functions in x, y, but can be immediately extended to the field Σ^* of meromorphic functions in x, y. If we put $x = x$, $y = x'y'$, every holomorphic function $\xi(x, y)$ assumes the form $x'^{\rho}(f_1(y') + x'f_2(y') + \cdots)$. If η is another holomorphic function of x, y, and $\eta = x'^{\sigma}(\phi_1(y') + x'\phi_2(y') + \cdots)$, we map the function ξ/η upon 0, ∞ or $f_1(y')/\phi_1(y')$, according as $\rho > \sigma$, $\rho < \sigma$ or $\rho = \sigma$. This mapping defines an homomorphism of the field Σ^* upon the purely transcendental field $\mathfrak{k}(y')$. We regard this homomorphism as a divisor of the second kind of Σ^*, since the prime ideal in $\mathfrak{k}\{x, y\}$ determined by this divisor is the 0-dimensional ideal $\mathfrak{p}^* = (x, y)$. We associate this divisor with the simple ideal $\mathfrak{p}^*$. In the same manner we may associate with any simple ideal $\mathfrak{P}_{h+1}$ of kind $h + 1$ in $\mathfrak{k}\{x, y\}$ a divisor of Σ^*, in which Σ^* is mapped upon a purely transcendental field of one variable. We do this by first applying h successive quadratic transformations, getting a ring $\mathfrak{O}^*_h = \mathfrak{k}\{x_h, y_h\}$, which contains $\mathfrak{k}\{x, y\}$ and in which the ideal $\mathfrak{P}_{h+1}$ corresponds to a simple v-ideal of kind 1, i. e. to the ideal (x_h, y_h). This ideal defines a divisor of the field Σ^*_h of meromorphic functions of x_h, y_h, mapping Σ^*_h upon the field $\mathfrak{k}(y'_h)$ of rational functions of y'_h $(= y_h/x_h)$. In this homomorphism the subfield Σ^* of Σ^*_h is mapped upon the entire field $\mathfrak{k}(y'_h)$, since x_h, y_h are rational functions of x and y. We associate the divisor of Σ^*, obtained in this manner, with the simple v-ideal $\mathfrak{P}_{h+1}$. In exactly the same manner as for the field of rational functions of x and y, it is shown that the correspondence between all the divisors of Σ^*, defined by homomorphic mappings of Σ upon fields of dimension one with respect to $\mathfrak{k}$ and of second kind with respect to $\mathfrak{k}\{x, y\}$, and the simple v-ideals in $\mathfrak{k}\{x, y\}$ is $(1, 1)$, and that consequently any such divisor is purely transcendental (i. e. the field upon which Σ^* is mapped by a divisor of the second kind is necessarily purely transcendental).

We point out explicitly, that if a divisor of Σ^* *of the first kind with respect to* $\mathfrak{k}\{x, y\}$ is defined as an homomorphic mapping of Σ^* upon a field Ω, such that the prime ideal determined in $\mathfrak{k}\{x, y\}$ by the divisor is one-dimensional, then Ω is necessarily the field of all meromorphic functions of one indeterminate. Thus for fields of meromorphic functions Σ^* the classification of the divisors into two kinds is not merely a relative classification with respect to the ring $\mathfrak{k}\{x, y\}$, but rather a classification in terms of the properties of the field Σ^* itself. That this should be so is only natural, in view of the privileged rôle which the ring of holomorphic functions plays in the field Σ^*.

We conclude with one final remark, which finds an application in the proof of the well-known algebro-geometric theorem, that a pencil of curves on an algebraic surface is necessarily linear, if the pencil has a base point at a simple point of the surface. This remark will be elaborated in a joint note by Dr. O. Schilling and the present author. At this place we wish only to observe that the proof of this theorem is based upon the following assertion:

Given a meromorphic function $f(x,y)/\phi(x,y)$, *and assuming that the elements* f *and* ϕ *of* $\mathfrak{k}\{x,y\}$ *are relatively prime and that neither is a unit, then there exists a divisor* $\mathfrak{P}$ *of second kind of the field* Σ^* *of meromorphic functions, such that* f/ϕ *is mapped upon a transcendental element of the image field.* This assertion can be readily proved as follows. Consider the ideal $\mathfrak{A} = (f, \phi)$ in $\mathfrak{k}\{x,y\}$. This ideal is zero-dimensional, since f and ϕ are relatively prime and are both contained in the ideal (x,y). Let $\mathfrak{A}'$ be the complete ideal determined by $\mathfrak{A}$. Let us first consider the case in which $\mathfrak{A}' = \mathfrak{p}^* = (x,y)$. In this case we assert that the required divisor $\mathfrak{P}$ is the one associated with $\mathfrak{p}^*$. In fact, assume that f/ϕ is mapped by $\mathfrak{P}$ upon a constant c. We may assume $c = 0$, replacing $f - c\phi$ by f. Under this hypothesis f/ϕ will have positive value in all the zero-dimensional valuations B_α defined by the points of the Riemann surface of $\mathfrak{P}$, and hence, since $\phi \equiv 0(\mathfrak{p}^*)$, f must belong, for any α, to the valuation ideal $\mathfrak{q}_2^{(\alpha)}$ which follows $\mathfrak{p}^*$ in the Jordan sequence of v-ideals belonging to B_α. As α varies, $\mathfrak{q}_2^{(\alpha)}$ can be any maximal subideal of $\mathfrak{p}^*$ (see section 8) and is a simple ideal (of kind 2). Now for *some* α we will have $\phi \equiv 0(\mathfrak{q}_2^\alpha)$. In fact, it is sufficient to consider the valuation defined by an irreducible branch of the analytical locus $\phi = 0$. For this valuation it is true that the element ϕ is contained in *all* the 0-dimensional v-ideals $\mathfrak{q}_i$ belonging to the valuation, whence also in $\mathfrak{q}_2$. It follows then that for some α, both f and ϕ are contained in $\mathfrak{q}_2^{(\alpha)}$. But this is impossible, since $\mathfrak{q}_2^{(\alpha)}$ is a complete ideal which does not contain the complete ideal $\mathfrak{A}' = \mathfrak{p}^*$ determined by (f, ϕ).

In the general case, let $\mathfrak{A}' = \mathcal{P}_1^{a_1}\mathcal{P}_2^{a_2}\cdots\mathcal{P}_k^{a_k}$, where $\mathcal{P}_i$ is a simple v-ideal of kind h_i. Let $h = \max\{h_i\}$, and let us apply the quadratic transformation $x' = x$, $y' = y/x$. If $\mathfrak{A} \equiv 0(\mathfrak{p}^{*\rho})$, $\mathfrak{A} \not\equiv 0(\mathfrak{p}^{*\rho+1})$, then by Theorem 12.3, also $\mathfrak{A}' \equiv 0(\mathfrak{p}^{*\rho})$, $\mathfrak{A}' \not\equiv 0(\mathfrak{p}^{*\rho+1})$, and putting $f = x'^\rho f_1$, $\phi = x'^\rho \phi_1$, $\mathfrak{A}_1 = (f_1, \phi_1)$, the complete ideal $\mathfrak{A}'_1$ determined by $\mathfrak{A}_1$ is the transform $T(\mathfrak{A}')$, i. e. the ideal

$$\mathfrak{A}'_1 = \mathcal{P}'^{a_1}_1\mathcal{P}'^{a_2}_2\cdots\mathcal{P}'^{a_k}_k.$$

Here each $\mathcal{P}'_i$ is of kind $h_i - 1$, so that $\max\{h_i - 1\} = h - 1$. Since the theorem has already been proved for $h = 1$ ($\mathfrak{A}' = \mathfrak{p}^*$), we may assume that the theorem is true for $h - 1$. We may even assume that for the function f_1/ϕ_1 the divisor whose existence is stated in our assertion is the divisor associated with a factor $\mathcal{P}'_s$, such that $\mathcal{P}'_s$ is of maximum kind $h - 1$. Then it follows immediately that the divisor of the field Σ^* associated with the factor $\mathcal{P}_s$ satisfies the assertion.

THE JOHNS HOPKINS UNIVERSITY.

INTEGRAL FORMS AND VARIATIONAL ORTHOGONALITY.*

By PHILIP HARTMAN and RICHARD KERSHNER.

Introduction. The arc-length of a curve $x_1 = x_1(t)$; $x_2 = x_2(t)$, if it exists, is defined [1] by an integral of the form

$$L(x_1, x_2) = \int_a^b \{(dx_1)^2 + (dx_2)^2\}^{\frac{1}{2}}, \tag{1}$$

where (1) means the limit of approximating sums

$$\Sigma\{\Delta x_1)^2 + (\Delta x_2)^2\}^{\frac{1}{2}}$$

of the Riemann-Stieltjes type. It is well known that (1) exists whenever x_1 and x_2 are of bounded variation in $[a, b]$; and that, in case x_1 and x_2 are absolutely continuous, (1) reduces to the ordinary Lebesgue integral

$$L(x_1, x_2) = \int_a^b \{(x'_1(t))^2 + (x'_2(t))^2\}^{\frac{1}{2}} dt. \tag{2}$$

In this paper the integral (1) will be investigated in cases when x_1 and x_2 are not absolutely continuous but possess, under the Lebesgue decomposition, a purely singular or a purely discontinuous component. The results to be obtained furnish means for the calculation of the length of curves given by a parametric representation for which (2) is not valid. These results will be based on the fact to be proved that, while (1) is clearly not decomposed linearly under arbitrary linear decompositions of x_1, x_2, nevertheless, for the Lebesgue decomposition, $x_i(t) = a_i(t) + s_i(t) + p_i(t)$, where a_i is absolutely continuous, s_i is purely singular,[2] and p_i is purely discontinuous, one has

$$L(x_1, x_2) = L(a_1, a_2) + L(s_1, s_2) + L(p_1, p_2). \tag{3}$$

It should be noticed that, since $L(a_1, a_2)$ reduces to an ordinary Lebesgue integral and $L(p_1, p_2)$ degenerates into an infinite sum, the usefulness of (3) is limited largely by the difficulty of calculating $L(s_1, s_2)$. Of course, in special cases it may be possible to eliminate the parameter between $s_1(t)$ and $s_2(t)$ and reduce this integral to one of the Lebesgue type. Additional results which sometimes lead to the explicit evaluation of $L(s_1, s_2)$ will be given in

---

* Received October 1, 1937.

[1] Cf., e. g., Saks [7], p. 57.

[2] A purely singular function will be supposed to be continuous.

the sequel. In this direction the following fact may be mentioned here: If, for every t in $[a, b]$, either $s'_1(t) = 0$ or $s'_2(t) = 0$, then

$$L(s_1, s_2) = V\{[a, b]; s_1\} + V\{[a, b]; s_2\},$$

where $V\{[a, b]; f\}$ is the total variation of f on $[a, b]$.

In the non-parametric case, the length $L(y)$ of the continuous curve $y = y(x)$ is given by

$$L(y) = \int_a^b \{1 + (y'(t))^2\}^{\frac{1}{2}} dt + V\{[a, b]; y\} - \int_a^b | y'(t) | dt.$$

In particular, if $y(x)$ is purely singular and monotone in $[a, b]$,

$$L(y) = | y(b) - y(a) | + | b - a |.$$

This last statement is trivial in the case that $y(x)$ is almost everywhere constant.

At the suggestion of Professor Wintner the proofs of the facts mentioned above have been extended so as to apply, not only to the Euclidean arc-length (1), but to very general integral forms of the type

$$\Phi(\mathbf{X}) = \Phi\{[a, b); \mathbf{X}\} = \int_a^b F(\mathbf{Y}; d\mathbf{X}), \tag{4}$$

where

$$\mathbf{Y} = \mathbf{Y}(t) = \{y_1(t), \cdots, y_m(t)\} \quad \text{and} \quad \mathbf{X} = \mathbf{X}(t) = \{x_1(t), \cdots, x_n(t)\}$$

are vector functions of t on $[a, b]$ and where

$$F(\mathbf{Y}; k\mathbf{Z}) = kF(\mathbf{Y}; \mathbf{Z}), \qquad (k \geqq 0).$$

The best known integrals of this type are, of course, the Stieltjes integral

$$\Phi(\mathbf{X}) = \int_a^b y(t)\, dx(t) \tag{5}$$

and the total variation of the function $x(t)$

$$\Phi(\mathbf{X}) = V\{[a, b); x\} = \int_a^b | dx |. \tag{6}$$

Another simple example is the case

$$\Phi(\mathbf{X}) = \int_a^b \{\sum_{j=1}^{n} | dx_j |^p\}^{1/p}, \qquad (p > 1), \tag{7}$$

of which the case $p = 2$ is Euclidean arc-length. The Riemannian arc-length

$$\Phi(\mathbf{X}) = \int_a^b \{\sum_{i,k=1}^{n} g_{ik} dx_i dx_k\}^{\frac{1}{2}} \tag{8}$$

and the Finsler metrics

$$\Phi(\boldsymbol{X}) = \int_a^b F(\boldsymbol{X}; d\boldsymbol{X}) \tag{9}$$

can also be considered as special cases of (4).

In connection with Hilbert's theory of bounded quadratic forms with continuous spectra, Hellinger [4] and Hahn [2] consider integrals of the form

$$\Phi(\boldsymbol{X}) = H_1(x_1, x_2) = \int_a^b (dx_1)^2/dx_2 \tag{10}$$

and

$$\Phi(\boldsymbol{X}) = H_2(x_1, x_2) = \int_a^b | dx_1 dx_2 |^{\frac{1}{2}}, \tag{11}$$

and need, in order to describe the system of orthogonal invariants, especially a study of the case that $\boldsymbol{X}(t)$ is not absolutely continuous. The general results to be obtained provide an essential simplification of the use of these integrals in this connection.

In general, it is found, in conformity with (3), that if

$$\boldsymbol{X}(t) = \boldsymbol{X}_1(t) + \boldsymbol{X}_2(t) + \boldsymbol{X}_3(t), \tag{12}$$

where $\boldsymbol{X}(t)$ is of bounded variation, $\boldsymbol{X}_1(t)$ is absolutely continuous, $\boldsymbol{X}_2(t)$ is purely singular, and $\boldsymbol{X}_3(t)$ is purely discontinuous, then, under very general conditions on the function F occurring in (4),

$$\Phi(\boldsymbol{X}) = \Phi(\boldsymbol{X}_1) + \Phi(\boldsymbol{X}_2) + \Phi(\boldsymbol{X}_3). \tag{13}$$

In particular (13) holds when $\Phi(\boldsymbol{X})$ is any one of the explicit types (5), (6), (7), (8), (10), (11). In the case of (9) some restrictions on F are, of course, required.

Actually the Lebesgue decomposition (12) is only a special case of a large class of decompositions for which the linear property exhibited in (13) is demonstrated. Other decompositions, of a similar nature, are considered for which (13) does not hold precisely but requires corrective terms.

Again in conformity with the case of Euclidean arc-length, it is shown that the first integral $\Phi(\boldsymbol{X}_1)$ occurring on the right side of (13) can be reduced, in all cases for which (13) is proved, to an ordinary Lebesgue integral with respect to t, namely

$$\Phi(\boldsymbol{X}_1) = \int_a^b F(\boldsymbol{Y}(t); \boldsymbol{X}'(t))\,dt, \qquad (\boldsymbol{X}' = d\boldsymbol{X}/dt).$$

Furthermore, the last integral, $\Phi(\boldsymbol{X}_3)$, occurring on the right of (13) degenerates to an ordinary series of the form

$$\Phi(\boldsymbol{X}_3) = \Sigma F(\boldsymbol{Y}(\rho_i); \boldsymbol{X}(\rho_i + 0) - \boldsymbol{X}(\rho_i - 0)),$$

where the sum is taken over all discontinuity points ρ_i of $\boldsymbol{X}(t)$. Thus (13), in a sense, reduces the investigation of integrals of type (4) to the case that $\boldsymbol{X}(t)$ is purely singular, i. e., that all components of $\boldsymbol{X}(t)$ are purely singular.

The results to be obtained are, of course, trivial in the Stieltjes case (5) and essentially known in the case (6) of total variation but in all other cases they seem to provide a new analysis of the integrals under consideration.

1. Definitions and conditions. The integral (4) may be defined as a direct generalization of the Riemann-Stieltjes integral in the following manner: Let $a = t_0 < t_1 < \cdots < t_{p+1} = b$ be a mesh on the interval $[a, b]$. Then $\Phi(\boldsymbol{X})$ is the limit of an arbitrary sequence of sums

$$(14) \quad \sum_{j=0}^{p} F(\boldsymbol{Y}(\xi_j); \Delta_j \boldsymbol{X}), \quad (\Delta_j \boldsymbol{X} = \boldsymbol{X}(t_{j+1}) - \boldsymbol{X}(t_j); \ t_j \leqq \xi_j < t_{j+1}),$$

associated with any sequence of meshes for which $\Delta = \max | t_{j+1} - t_j | \rightarrow 0$, provided this limit exists and is independent of the particular choice of the sequence of meshes and of the intermediary values ξ_j. Correspondingly, the integral (4) is said to exist if the sums (14) have such a unique limit. The integral (4) will be said to exist absolutely if

$$(15) \qquad \int_a^b | F(\boldsymbol{Y}; d\boldsymbol{X}) |$$

as well as (4), exists. The corresponding generalization of the Lebesgue-Stieltjes type will not be considered here.

It will be convenient to list certain of the conditions which will be used in the sequel.

Throughout the paper it will be supposed that $\boldsymbol{X}(t)$ is of bounded variation, i. e., each component $x_j(t)$ is of bounded variation, in $[a, b]$. It will also be supposed that $\boldsymbol{X}(t)$ is continuous from the left. This involves no loss of generality, since it necessitates, at most, the redefinition of $\boldsymbol{X}(t)$ at an enumerable set of points, a process which does not affect the value of the integral (4).

No restrictions will be imposed on $\boldsymbol{Y}(t)$ other than that its components be Baire functions. Even this requirement is unnecessary for many purposes.

Of $F(\boldsymbol{Y}; \boldsymbol{Z})$, in addition to

(A) Positive linear homogeneity: $F(\boldsymbol{Y}; k\boldsymbol{Z}) = kF(\boldsymbol{Y}; \boldsymbol{Z})$, $(k \geqq 0)$, which has already been mentioned, it will always be supposed that F is continuous in the n variables z_j together.

Conditions which will be required not always but on occasion, are the following:

(B) Polar symmetry: $F(\boldsymbol{Y}(t); -\boldsymbol{Z}) = F(\boldsymbol{Y}(t); \boldsymbol{Z})$.

(C) Convexity: $F(\boldsymbol{Y}(t); \boldsymbol{Z}_1 + \boldsymbol{Z}_2) \leqq F(\boldsymbol{Y}(t); \boldsymbol{Z}_1) + F(\boldsymbol{Y}(t); \boldsymbol{Z}_2)$. In virtue of (A), for $k = \frac{1}{2}$, this condition is equivalent to the usual notion of convexity with respect to $\boldsymbol{Z}$.

(D) For every $\epsilon > 0$, one can choose a $\delta_\epsilon > 0$ such that

$$\left| \sum_{j=1}^{\infty} F(\boldsymbol{Y}(\xi_j); \boldsymbol{Z}_j) \right| < \epsilon \quad \text{if} \quad \sum_{j=1}^{\infty} \sum_{k=1}^{n} | z_{jk} | < \delta_\epsilon,$$

where $\boldsymbol{Z}_j = \{z_{j1}, \cdots, z_{jn}\}$ and $a \leqq \xi_j \leqq b$.

(D′) If $\boldsymbol{Y}(t)$ and $\boldsymbol{X}(t)$ are vector functions for which (4) exists, then for every $\epsilon > 0$, there exists a $\delta_\epsilon > 0$ such that

$$\left| \sum_{j=1}^{\infty} F(\boldsymbol{Y}(\xi_j); \Delta_j \boldsymbol{X}) \right| < \epsilon \quad \text{if} \quad \sum_{j=1}^{\infty} \sum_{k=1}^{n} | \Delta_j x_k | < \delta_\epsilon,$$

where $\{[t_j, t'_j)\}$ is any set of non-overlapping, half-open intervals on $[a, b]$, $\Delta_j x_k = x_k(t'_j) - x_k(t_j)$, and $t_j \leqq \xi_j < t'_j$.

(E) For every $\epsilon > 0$, there exists a $\delta_\epsilon > 0$ such that

$$\left| \sum_{j=1}^{\infty} \{ F(\boldsymbol{Y}(\xi_j); \boldsymbol{Z}_j) - F(\boldsymbol{Y}(\xi_j); \boldsymbol{W}_j) \} \right| < \epsilon \quad \text{if} \quad \sum_{j=1}^{\infty} \sum_{k=1}^{n} | z_{jk} - w_{jk} | < \delta_\epsilon,$$

where $\boldsymbol{Z}_j = \{z_{j1}, \cdots, z_{jn}\}$, $\boldsymbol{W}_j = \{w_{j1}, \cdots, w_{jn}\}$, and $a \leqq \xi_j \leqq b$.

(E′) If $\boldsymbol{Y}(t), \boldsymbol{X}_1(t), \boldsymbol{X}_2(t)$ are vector functions for which $\int_a^b F(\boldsymbol{Y}; d\boldsymbol{X}_1)$ and $\int_a^b F(\boldsymbol{Y}; d\boldsymbol{X}_2)$ exist, then one can choose, for every $\epsilon > 0$, a $\delta_\epsilon > 0$ such that

$$\left| \sum_{j=1}^{\infty} \{ F(\boldsymbol{Y}(\xi_j); \Delta_j \boldsymbol{X}_1) - F(\boldsymbol{Y}(\xi_j); \Delta_j \boldsymbol{X}_2) \} \right| < \epsilon$$
$$\text{if} \sum_{j=1}^{\infty} \sum_{k=1}^{n} | \Delta_j x_{1k} - \Delta_j x_{2k} | < \delta_\epsilon,$$

the notations being analogous to those used in (D′).

The relations between these last four conditions are indicated by the following table of implications,

$$(\text{E}) \cdot\supset\cdot (\text{E}'), \quad (\text{D}) \cdot\supset\cdot (\text{D}'), \quad (\text{E}) \cdot\supset\cdot (\text{D}), \quad (\text{E}') \cdot\supset\cdot (\text{D}'),$$

where, e. g., (E) $\cdot\supset\cdot$ (E′) means that (E′) is satisfied whenever (E) is. Condition (D) is somewhat weaker than a Lipschitz condition on $F(\boldsymbol{Y}(t); \boldsymbol{Z})$

with respect to the n variables z_j at the point $z_j = 0$, which is uniform with respect to the parameter t in the t-interval $[a, b]$. Correspondingly, (E) is satisfied if $F(\mathbf{Y}(t); \mathbf{Z})$ fulfills a uniform Lipschitz condition in the n independent variables z_j.

2. $\Phi(\mathbf{X})$ as a set function. For later purposes it will be convenient to extend the definition of the integral $\Phi(\mathbf{X}) = \Phi\{[a, b); \mathbf{X}\}$ to a set function $\Phi\{S; \mathbf{X}\}$, defined for all Borel sets S in $[a, b]$. It is known[3] that if the point function $f(u; \mathbf{X}) = \Phi\{[a, u); \mathbf{X}\}$ exists for every u in $a < u \leqq b$, is of bounded variation, and continuous from the left in $a < u \leqq b$, then there exists a unique, bounded, completely additive set function $\Phi\{S; \mathbf{X}\}$, defined for all Borel sets S in $[a, b]$, such that if S is the half-open interval $[a, u)$, $a \leqq t < u$, then $\Phi\{S; \mathbf{X}\} = f(u; \mathbf{X})$.

This set function $\Phi\{S; \mathbf{X}\}$ may be obtained in the following way: Let

$$f(u; \mathbf{X}) = f_1(u; \mathbf{X}) - f_2(u; \mathbf{X}),$$

where f_1 and f_2 are non-decreasing functions of u in $[a, b]$ and let

$$\Phi_i\{S; \mathbf{X}\} = \text{g. l. b.} \sum_k [f_i(u'_k) - f_i(u_k)], \qquad (i = 1, 2),$$

where $a \leqq u_k < u'_k \leqq b$ and $S \subset \sum_k [u_k, u'_k)$. Then

$$\Phi\{S; \mathbf{X}\} = \Phi_1\{S; \mathbf{X}\} - \Phi_2\{S; \mathbf{X}\}.$$

When such an extension is possible, the notation

$$\Phi\{S; \mathbf{X}\} = \int_S F(\mathbf{Y}; d\mathbf{X}) \tag{16}$$

will be used, in conformity with (4).

Now there will be proved

THEOREM I. *If* (4) *exists absolutely and F satisfies condition* (D'), *then*

$$f(u; \mathbf{X}) = \int_a^u F(\mathbf{Y}; d\mathbf{X}) \tag{17}$$

exists for every u in $a < u \leqq b$ and there exists a completely additive set function (16) *defined for all Borel sets in* $[a, b]$ *such that*

$$\Phi\{[a, u); \mathbf{X}\} = \int_a^u F(\mathbf{Y}; d\mathbf{X}).$$

Proof. The fact that (17) exists for every u in $[a, b]$ is, of course, a consequence of the existence of $f(b;\mathbf{X})$. In fact, for every $\epsilon > 0$, one can choose a $\Delta_\epsilon > 0$ such that if

---

[3] Cf., e. g., Radon [6], pp. 1305-1313.

$$a = t_0 < t_1 < \cdots < t_{p+1} = b \tag{18}$$

is any mesh for which the degree of fineness Δ satisfies

$$\Delta = \max | t_{j+1} - t_j | < \Delta_\epsilon, \tag{19}$$

then

$$\left| \sum_{j=0}^{p} F(\mathbf{Y}(\xi_j)\,;\, \Delta_j\mathbf{X}) - \int_a^b F(\mathbf{Y}; d\mathbf{X}) \right| < \epsilon \tag{20}$$

for arbitrary intermediary values ξ_j. Now if

$$a = t_0 < t_1 < \cdots < t_{k+1} = u < t_{k+2} < \cdots < t_{p+1} = b$$

and

$$a = t'_0 < t'_1 < \cdots < t'_{r+1} = u < t'_{r+2} = t_{k+2} < t'_{r+3} = t_{k+3} < \cdots < t'_{s+1} = b$$

are any two meshes having degrees of fineness Δ, Δ' respectively, which satisfy (19), then

$$\left| \sum_{j=0}^{k} F(\mathbf{Y}(\xi_j)\,;\, \Delta_j\mathbf{X}) + \sum_{j=k+1}^{p} F(\mathbf{Y}(\xi_j)\,;\, \Delta_j\mathbf{X}) - \int_a^u F(\mathbf{Y}; d\mathbf{X}) \right| < \epsilon,$$

$$\left| \sum_{j=0}^{r} F(\mathbf{Y}(\xi'_j)\,;\, \Delta'_j\mathbf{X}) + \sum_{j=k+1}^{p} F(\mathbf{Y}(\xi_j)\,;\, \Delta_j\mathbf{X}) - \int_a^b F(\mathbf{Y}; d\mathbf{X}) \right| < \epsilon.$$

Consequently,

$$\left| \sum_{j=0}^{k} F(\mathbf{Y}(\xi_j)\,;\, \Delta_j\mathbf{X}) - \sum_{j=0}^{r} F(\mathbf{Y}(\xi'_j)\,;\, \Delta'_j\mathbf{X}) \right| < 2\epsilon \tag{21}$$

whenever $\Delta < \Delta_\epsilon$ and $\Delta' < \Delta_\epsilon$. Thus (17) exists for every u in $[a, b]$.

As was mentioned at the beginning of this section, in order to complete the proof of Theorem I, it is sufficient to show that the function (17) is of bounded variation and continuous from the left on $a < u \leqq b$. That $f(u;\mathbf{X})$ is of bounded variation may be seen from the decomposition

$$f(u;\mathbf{X}) = \int_a^u | F(\mathbf{Y}; d\mathbf{X}) | - \int_a^u \{ | F(\mathbf{Y}; d\mathbf{X}) | - F(\mathbf{Y}; d\mathbf{X}) \}$$

of f into the difference of two monotone functions. The existence of the integrals on the right side is assured by the assumption of the existence of (4) and (15), as has been shown above.

Since $\mathbf{X}(t)$ is of bounded variation and continuous from the left, for every u, $a < u \leqq b$, there exists an $\eta = \eta_{\epsilon,u}$ such that the variation

$$V\{[u', u)\,;\, \mathbf{X}\} = \int_{u'}^{u} \sum_{k=1}^{n} | dx_k | < \delta_\epsilon \text{ if } 0 < u - u' < \eta,$$

where δ_ϵ is defined in (D′). In virtue of (D′), it is seen for any mesh $u' = t_0 < t_1 < \cdots < t_{p+1} = u$, that

$$\left| \sum_{j=0}^{p} F(\mathbf{Y}(\xi_j)\,;\,\Delta_j\mathbf{X}) \right| < \epsilon, \text{ since } \sum_{j=0}^{p}\sum_{k=1}^{n} |\Delta_j x_k| \leqq V\{[u',u)\,;\,X\} < \delta_\epsilon.$$

This implies that

$$| f(u;\mathbf{X}) - f(u';\mathbf{X}) | = \left| \int_{u'}^{u} F(\mathbf{Y};d\mathbf{X}) \right| < \epsilon \text{ if } 0 < u - u' < \eta,$$

completing the proof of Theorem I.

Obvious modifications of the proof of the last part of Theorem I show that if S is any Borel set for which $V\{S;\mathbf{X}\} = 0$, then $\Phi\{S;\mathbf{X}\} = 0$. In particular one has

THEOREM II. *Under the assumptions of Theorem I, the functions* $f(u;\mathbf{X})$, $\Phi\{S;\mathbf{X}\}$ *are absolutely continuous, purely singular, or purely discontinuous with* $\mathbf{X}(t)$.

Similarly, the first part of Theorem I can easily be extended to show that if S is a set consisting of a finite number of disjoint half-open intervals $[u_j, u'_j)$, $j = 1, \cdots, s$, and if

$$u_j = t_{j0} < t_{j1} < \cdots < t_{jp_{j+1}} = u'_j$$

is a mesh over S such that the degree of fineness Δ satisfies $\Delta < \Delta_\epsilon$, then

$$\left| \sum_{j=1}^{s}\sum_{k=0}^{p_j} F(\mathbf{Y}(\xi_{j,k})\,;\,\mathbf{X}(t_{j,k+1}) - \mathbf{X}(t_{j,k})) - \int_S F(\mathbf{Y};d\mathbf{X}) \right| < 2\epsilon.$$

Use of this fact will be made in the proof of Theorem III.

3. Reduction to other forms. It has been shown that in case $\mathbf{X}(t)$ is absolutely continuous, the integral (4) reduces in many cases to an ordinary Lebesgue integral. The precise formulation of the result to be proved in this direction is

THEOREM III. *If the assumptions of Theorem I are satisfied and if* $\mathbf{X}(t)$ *is absolutely continuous, then*

$$\int_a^b F(\mathbf{Y};d\mathbf{X}) = \int_a^b F(\mathbf{Y}(t)\,;\,\mathbf{X}'(t))\,dt. \tag{22}$$

For the proof of this theorem, the following lemma will be needed.

LEMMA 1. *If the assumptions of Theorem III are satisfied, then for any Borel set* T *on which* $\mathbf{X}'(t)$ *exists and*

$$\lambda \leqq F(\mathbf{Y}(t);\ \mathbf{X}'(t)) < \mu \tag{23}$$

one has

$$\lambda \operatorname{meas} T \leqq \Phi\{T; \mathbf{X}\} \leqq \mu \operatorname{meas} T. \tag{24}$$

Proof. By Theorem II, the function $\Phi\{S; \mathbf{X}\}$ is absolutely continuous, i. e., for every $\epsilon > 0$, one can choose an $\eta_\epsilon > 0$ such that $|\Phi\{S; \mathbf{X}\}| < \epsilon$ when S is any Borel set for which meas $S < \eta_\epsilon$.

Now, since the set T occurring in the statement of the lemma is a bounded set, the covering theorem of Vitali implies that there exists a set $U = \sum_{j=1}^{p} [u_j, u'_j)$ such that

$$0 < u'_j - u_j < \Delta_\epsilon; \qquad u'_j \leqq u_{j+1}; \tag{25}$$

$$\operatorname{meas}(T - TU) < \eta_\epsilon; \qquad \operatorname{meas}(U - TU) < \eta_\epsilon; \tag{26}$$

$$|\operatorname{meas} T - \operatorname{meas} U| < \epsilon; \tag{27}$$

$$\lambda - \epsilon \leqq F(\mathbf{Y}(\xi_j);\ [\mathbf{X}(u'_j) - \mathbf{X}(u_j)]/[u'_j - u_j]) < \mu; \quad u_j \leqq \xi_j < u'_j. \tag{28}$$

Now, by (28) and the property (A) of F, one has

$$(\lambda - \epsilon)(u'_j - u_j) \leqq F(\mathbf{Y}(\xi_j);\ \mathbf{X}(u'_j) - \mathbf{X}(u_j)) < \mu(u'_j - u_j). \tag{29}$$

By adding the inequalities (29), one obtains

$$(\lambda - \epsilon) \operatorname{meas} U \leqq \sum_{j=1}^{p} F(\mathbf{Y}(\xi_j); \mathbf{X}(u'_j) - \mathbf{X}(u_j)) < \mu \operatorname{meas} U. \tag{30}$$

In virtue of (25) and the remark made immediately after Theorem II,

$$\left| \sum_{j=1}^{p} F(\mathbf{Y}(\xi_j);\ \mathbf{X}(u'_j) - \mathbf{X}(u_j)) - \int_U F(\mathbf{Y}; d\mathbf{X}) \right| < 2\epsilon. \tag{31}$$

Now

$$\left| \int_T - \int_U \right| = \left| \int_{T-TU} - \int_{U-TU} \right| \leqq \left| \int_{T-TU} \right| + \left| \int_{U-TU} \right| < 2\epsilon \tag{32}$$

by (26) and the definition of η_ϵ. Finally, by (27), (30), (31), and (32),

$$(\lambda - \epsilon)(\operatorname{meas} T - \epsilon) - 4\epsilon \leqq \Phi\{T; \mathbf{X}\} < \mu(\operatorname{meas} T + \epsilon) + 4\epsilon.$$

The inequality (24) now follows since $\epsilon > 0$ is arbitrary.

Proof of Theorem III. Consider a lower and an upper Lebesgue approximating sum for the function $F(\mathbf{Y}(t); \mathbf{X}'(t))$. Then, by Lemma 1,

$$\sum_{i=-\infty}^{+\infty} \lambda_i \operatorname{meas} S\{\lambda_i \leqq F < \lambda_{i+1}\} \leqq \sum_{i=-\infty}^{+\infty} \Phi\{S_i; \mathbf{X}\} = \Phi\{[a, b); \mathbf{X}\} \tag{33}$$

and

$$(34) \qquad \sum_{i=-\infty}^{+\infty} \lambda_i \operatorname{meas} S\{\lambda_{i-1} < F \leqq \lambda_i\} \geqq \sum_{i=-\infty}^{+\infty} \Phi\{S'_i; \mathbf{X}\} = \Phi\{[a,b); \mathbf{X}\}$$

where $S_i = S\{\lambda_i \leqq F < \lambda_{i+1}\}$ is the set of points t in $[a,b]$ for which $\lambda_i \leqq F(Y(t); X'(t)) < \lambda_{i+1}$ and $S'_i = S\{\lambda_{i-1} < F \leqq \lambda_i\}$ has an analogous meaning. The inequalities (33) and (34) imply that the integral occurring on the right of (22) exists and that (22) holds.

By Theorem II, if $\mathbf{X}(t)$ is purely discontinuous then $\Phi\{S; \mathbf{X}\}$ is a purely discontinuous set function. In this case, also, the integral has a simple form.

THEOREM IV. *If the assumptions of Theorem I are satisfied and if* $\mathbf{X}(t)$ *is purely discontinuous, then*

$$\int_a^b F(Y; d\mathbf{X}) = \sum F(\mathbf{Y}(\rho_i); \mathbf{X}(\rho_i + 0) - \mathbf{X}(\rho_i - 0)),$$

where the sum extends over all discontinuity points ρ_i *of* $\mathbf{X}(t)$.

Proof. By the remark immediately following the proof of Theorem I, it is seen that

$$(35) \qquad \int_a^b F(\mathbf{Y}; d\mathbf{X}) = \int_{\Sigma\rho_i} F(\mathbf{Y}; d\mathbf{X}) = \sum \int_{\rho_i} F(\mathbf{Y}; d\mathbf{X}).$$

But, by definition,

$$(36) \qquad \int_{\rho_i} F(\mathbf{Y}; d\mathbf{X}) = \lim_{\epsilon \to 0} \int_{\rho_i - \epsilon}^{\rho_i + \epsilon} F(\mathbf{Y}; d\mathbf{X})$$
$$= \lim_{\epsilon \to 0} F(\mathbf{Y}(\rho_i); \mathbf{X}(\rho_i + \epsilon) - \mathbf{X}(\rho_i - \epsilon)).$$

Theorem IV is a consequence of these relations.

4. Variational orthogonality. Two functions $x_1(t)$, $x_2(t)$ of bounded variation on $a \leqq t \leqq b$ will be said to be variationally orthogonal on $[a,b]$ if there exists a Borel set S such that

$$(37_1) \qquad V\{S; x_1\} = V\{[a,b]; x_1\}; \qquad (37_2) \qquad V\{S; x_2\} = 0,$$

where, as above, $V\{S; x\} = \int_S |dx|$ is the total variation of the function x on the set S. (It is clear that the seeming lack of symmetry of this definition with respect to x_1 and x_2 is only apparent.) For example, if $x_1(t)$ is purely singular and if $x_2(t)$ is absolutely continuous, then x_1 and x_2 are variationally orthogonal and S may be chosen to be a zero set satisfying (37_1). Similarly, if $x_1(t)$ is purely discontinuous and if $x_2(t)$ is continuous, then x_1 and x_2 are variationally orthogonal and S may be chosen to be the enumerable set of points of discontinuity of $x_1(t)$. Finally let $x(t)$ be any function of bounded variation and let

$$(38)\qquad x_1(t) = \int_a^t \tfrac{1}{2}\{|\,dx\,| + dx\}$$

be the positive variation of x on $[a, t)$ and

$$(39)\qquad x_2(t) = x_1(t) - x(t) = \int_a^t \tfrac{1}{2}\{|\,dx\,| - dx\},$$

the negative variation of x on $[a, t)$, so that,

$$(40)\qquad x(t) = x_1(t) - x_2(t)$$

is Jordan's decomposition of $x(t)$ into the difference of two non-decreasing functions. Then $x_1(t)$ and $x_2(t)$ are variationally orthogonal and S may be chosen to be the set where $x'(t) \geqq 0$ (including $x'(t) = +\infty$).

The two vector functions

$$(41)\quad \mathbf{X}_1(t) = \{x_{11}(t), \cdots, x_{1n}(t)\} \quad \text{and} \quad \mathbf{X}_2(t) = \{x_{21}(t), \cdots, x_{2n}(t)\}$$

will be said to be variationally orthogonal if $x_{1j}(t)$ and $x_{2j}(t)$ are variationally orthogonal for every j. The two vector functions (41) will be said to be completely variationally orthogonal if there exists a Borel set S for which

$$(42_1)\qquad V\{S;\mathbf{X}_1\} = V\{[a,b];\mathbf{X}_1\}; \qquad (42_2)\qquad V\{S;\mathbf{X}_2\} = 0,$$

i. e., for $j = 1, \cdots, n$,

$$(43_1)\qquad V\{S;x_{1j}\} = V\{[a,b];x_{1j}\}; \qquad (43_2)\qquad V\{S;x_{2j}\} = 0.$$

Thus, if $\mathbf{X}_1(t)$ is purely singular and $\mathbf{X}_2(t)$ is absolutely continuous, then $\mathbf{X}_1(t)$ and $\mathbf{X}_2(t)$ are completely variationally orthogonal and S may be chosen to be a zero set satisfying (42_1). The same remarks hold if $\mathbf{X}_1(t)$ is purely discontinuous and $\mathbf{X}_2(t)$ is continuous. However, if $\mathbf{X}(t)$ is an arbitrary vector function of bounded variation and if $\mathbf{X}(t) = \mathbf{X}_1(t) - \mathbf{X}_2(t)$, represents the decomposition (38), (39), (40) of its components into monotone parts, then $\mathbf{X}_1(t)$ and $\mathbf{X}_2(t)$ are variationally orthogonal but need not be completely so.

5. Linearity. The property of complete variational orthogonality discussed above is exactly what is needed to insure that the integral (4) satisfies the decomposition property exhibited in (13).

THEOREM V. *Suppose that* (4) *exists absolutely and that F satisfies condition* (E). *Let $\mathbf{X}(t) = \mathbf{X}_1(t) + \mathbf{X}_2(t)$, where $\mathbf{X}_1$ and $\mathbf{X}_2$ are completely variationally orthogonal. Then*

$$\int_a^b F(\mathbf{Y}; d\mathbf{X}) = \int_a^b F(\mathbf{Y}; d\mathbf{X}_1) + \int_a^b F(\mathbf{Y}; d\mathbf{X}_2), \tag{44}$$

the existence of the last two integrals being proved.

Proof. Since the conditions of Theorem I are satisfied, (16) exists and is a completely additive set function.

Let S denote the set satisfying (42). Let $\epsilon > 0$ be fixed. Let $U = U_\epsilon = \sum_{j=1}^{\infty} [u_j, u'_j)$ be an enumerable set of pairwise disjoint, half-open intervals which contains S. By the definition of $\Phi\{S; \mathbf{X}\}$, it is clear that U may be chosen so that

$$| \Phi\{S; \mathbf{X}\} - \Phi\{U_\epsilon; \mathbf{X}\} | < \epsilon; \tag{45}$$

while from (42_2),

$$V\{U_\epsilon; \mathbf{X}_2\} = \sum_{k=1}^{n} V\{U_\epsilon; x_{2k}\} < \delta_\epsilon, \tag{46}$$

where δ_ϵ is defined in E. Now, since Φ, V are completely additive set functions,

$$\Phi\{U_\epsilon; \mathbf{X}\} = \sum_{j=1}^{\infty} \Phi\{[u_j, u'_j);\ \mathbf{X}\}; \quad V\{U_\epsilon; \mathbf{X}\} = \sum_{j=1}^{\infty} V\{[u_j, u'_j);\ \mathbf{X}\}. \tag{47}$$

Thus, it is possible to select a sufficiently large integer $N = N_\epsilon$, such that

$$\left| \sum_{j=1}^{N} \Phi\{[u_j, u'_j);\ \mathbf{X}\} - \Phi\{U_\epsilon; \mathbf{X}\} \right| < \epsilon, \tag{48}$$

and at the same time,

$$\sum_{j=N+1}^{\infty} \sum_{k=1}^{n} V\{[u_j, u'_j);\ x_{1k}\} < \delta_\epsilon. \tag{49}$$

Let $a = t_0 < t_1 < \cdots < t_{p+1} = b$ be a mesh on $[a, b]$ such that each u_j and each u'_j, for $j = 1, \cdots, N$, occurs among the t_k. Let the degree of fineness Δ of this mesh be less than the number Δ_ϵ defined in (19) and (20). It is clear that the requirement that u_j and u'_j occur among the t_k is no essential restriction on the generality of this mesh. Let Σ'_k denote the sum over those k-values for which the interval $[t_k, t'_k)$ lies in an interval $[u_j, u'_j)$, $j \leqq N$. Let Σ''_k denote the sum over all remaining k-values. Then

$$\begin{aligned} &\left| \sum_{k=0}^{p} F(\mathbf{Y}(\xi_k);\ \Delta_k \mathbf{X}_1) - \Sigma'_k F(\mathbf{Y}(\xi_k);\ \Delta_k \mathbf{X}) \right| \\ &\quad \leqq | \Sigma'_k \{ F(\mathbf{Y}(\xi_k);\ \Delta_k \mathbf{X}_1) - F(\mathbf{Y}(\xi_k);\ \Delta_k \mathbf{X}) \} | + | \Sigma''_k F(\mathbf{Y}(\xi_k);\ \Delta_k \mathbf{X}_1) |. \end{aligned} \tag{50}$$

Now, by (46),

$$\Sigma'_k \sum_{j=1}^{n} | \Delta_k x_{1j} - \Delta_k x_j | = \Sigma'_k \sum_{j=1}^{n} | \Delta_k x_{2j} | \leqq \sum_{j=1}^{n} V\{U_\epsilon; x_{2j}\} < \delta_\epsilon, \tag{51}$$

and, by (49) and (42_1),

$$(52)\qquad \Sigma''_k \sum_{j=1}^{n} | \Delta_k x_{1j} | \leqq \sum_{j=N+1}^{\infty} \sum_{k=1}^{n} V\{[u_j, u'_j)\,;\ x_{1k}\} + V\{[a, b] - U\,; x_{1k}\} < \delta_\epsilon.$$

It follows, from (50), (51), (52) and the property (E) of F, that

$$(53)\qquad \left| \sum_{k=0}^{p} F(\mathbf{Y}(\xi_k)\,;\ \Delta_k\mathbf{X}_1) - \Sigma'_k F(\mathbf{Y}(\xi_k)\,;\ \Delta_k\mathbf{X}) \right| < 2\epsilon.$$

In virtue of the remark made following Theorem II and the fact that $\Delta < \Delta_\epsilon$,

$$(54)\qquad \left| \Sigma'_k F(\mathbf{Y}(\xi_k)\,;\ \Delta_k\mathbf{X}) - \sum_{j=1}^{N} \Phi\{u_j, u'_j)\,;\ \mathbf{X}\} \right| < 2\epsilon.$$

Addition of the inequalities (45), (48), (53), and (54) gives

$$(55)\qquad \left| \sum_{k=0}^{p} F(\mathbf{Y}(\xi_k)\,;\ \Delta_k\mathbf{X}_1) - \Phi\{S\,; \mathbf{X}\} \right| < 6\epsilon \text{ if } \Delta < \Delta_\epsilon.$$

Since the mesh employed is arbitrary, (55) shows that

$$(56)\qquad \int_a^b F(\mathbf{Y}\,; d\mathbf{X}_1) = \Phi\{S\,; \mathbf{X}\}$$

exists. In exactly the same way, it can be shown that

$$(57)\qquad \int_a^b F(\mathbf{Y}\,; d\mathbf{X}_2) = \Phi\{[a, b] - S\,; \mathbf{X}\}$$

exists. Since Φ is an additive set function, (56) and (57) together imply (44). This completes the proof of Theorem V.

The proof of Theorem V implies

THEOREM VI. *Suppose that* (4) *exists absolutely and that F satisfies condition* (E'). *Let $\mathbf{X}(t) = \mathbf{X}_1(t) + \mathbf{X}_2(t)$, where $\mathbf{X}_1$ and $\mathbf{X}_2$ are completely variationally orthogonal. Then*

$$\int_a^b F(\mathbf{Y}\,; d\mathbf{X}) = \int_a^b F(\mathbf{Y}\,; d\mathbf{X}_1) + \int_a^b F(\mathbf{Y}\,; d\mathbf{X}_2),$$

whenever the last two integrals exist.

In fact, condition (E) was used in the proof of Theorem V only in connection with the inequalities (50), (51), (52), (53) and it is obviously possible to use condition (E′) in the same way if the existence of $\int_a^b F(\mathbf{Y}\,; d\mathbf{X}_1)$ and $\int_a^b F(\mathbf{Y}\,; d\mathbf{X}_2)$ is assured.

THEOREM VII. *Suppose that* (4) *exists and that F satisfies conditions*

(B), (C), and (D). Let $\mathbf{X}(t)=\mathbf{X}_1(t)+\mathbf{X}_2(t)$, *where* $\mathbf{X}_1$ *and* $\mathbf{X}_2$ *are completely variationally orthogonal. Then*

$$\int_a^b F(\mathbf{Y};d\mathbf{X})=\int_a^b F(\mathbf{Y};d\mathbf{X}_1)+\int_a^b F(\mathbf{Y};d\mathbf{X}_2),$$

the existence of the last two integrals being proved.

Proof. In virtue of Theorem V, it is sufficient to prove that conditions (B), (C), and (D) imply condition (E) and the existence of (15). Now, by (C),

$$F(\mathbf{Y}(t);\,\mathbf{Z})-F(\mathbf{Y}(t);\,\mathbf{W})\leqq F(\mathbf{Y}(t);\,\mathbf{Z}-\mathbf{W})$$

and

$$F(\mathbf{Y}(t);\,\mathbf{W})-F(\mathbf{Y}(t);\,\mathbf{Z})\leqq F(\mathbf{Y}(t);\,\mathbf{W}-\mathbf{Z}).$$

Thus, by (B),

$$\begin{aligned}(58)\qquad &|F(\mathbf{Y}(t);\,\mathbf{Z})-F(\mathbf{Y}(t);\,\mathbf{W})|\\ &\qquad\leqq F(\mathbf{Y}(t);\,\mathbf{Z}-\mathbf{W})=F(\mathbf{Y}(t);\,\mathbf{W}-\mathbf{Z}).\end{aligned}$$

This inequality (58) shows, first of all, that $F(\mathbf{Y}(t);\,\mathbf{Z})\geqq 0$ for all $\mathbf{Z}$, so that existence coincides with absolute existence. Also (58) and the condition (D) imply (E).

THEOREM VIII. *Suppose that* (4) *exists and that* F *satisfies conditions (B), (C) and (D′). Let* $\mathbf{X}(t)=\mathbf{X}_1(t)+\mathbf{X}_2(t)$, *where* $\mathbf{X}_1$ *and* $\mathbf{X}_2$ *are completely variationally orthogonal. Then*

$$\int_a^b F(\mathbf{Y};d\mathbf{X})=\int_a^b F(\mathbf{Y};d\mathbf{X}_1)+\int_a^b F(\mathbf{Y};d\mathbf{X}_2),$$

whenever the last two integrals exist.

Proof. Examination of the proof of Theorems V and VI shows that condition (E′) is required only in the following forms,

$$\left|\sum_{j=1}^{\infty}\{F(\mathbf{Y}(\xi_j);\,\Delta_j\mathbf{X})-F(\mathbf{Y}(\xi_j);\,\Delta_j\mathbf{X}_1)\}\right|<\epsilon \text{ if } \sum_{j=1}^{\infty}\sum_{k=1}^{n}|\Delta_j x_{2k}|<\delta_\epsilon,$$

$$\left|\sum_{j=1}^{\infty}\{F(\mathbf{Y}(\xi_j);\,\Delta_j\mathbf{X})-F(\mathbf{Y}(\xi_j);\,\Delta_j\mathbf{X}_2)\}\right|<\epsilon \text{ if } \sum_{j=1}^{\infty}\sum_{k=1}^{n}|\Delta_j x_{1k}|<\delta_\epsilon.$$

Thus, the required formulae corresponding to (58) are

$$(59)\qquad \begin{aligned}&|F(\mathbf{Y}(\xi);\,\Delta\mathbf{X})-F(\mathbf{Y}(\xi);\,\Delta\mathbf{X}_1)|\leqq F(\mathbf{Y}(\xi);\,\Delta\mathbf{X}_2)\\ &|F(\mathbf{Y}(\xi);\,\Delta\mathbf{X})-F(\mathbf{Y}(\xi);\,\Delta\mathbf{X}_2)|\leqq F(\mathbf{Y}(\xi);\,\Delta\mathbf{X}_1).\end{aligned}$$

These formulae (59) are, obviously, consequences of (B) and (C). But (59) shows that the form of (E′) which is actually used in the proof of Theorem VI is implied by (D′). Thus Theorem VIII is a consequence of the proof of Theorem VI.

COROLLARY 1. *Let* $X(t) = X_1(t) + X_2(t) + X_3(t)$, *where* $X_1(t)$ *is absolutely continuous,* $X_2(t)$ *is purely singular, and* $X_3(t)$ *is purely discontinuous. Then, under the assumptions of any of Theorems V-VIII,*

$$(60)\quad \int_a^b F(Y; dX) = \int_a^b F(Y; dX_1) + \int_a^b F(Y; dX_2) + \int_a^b F(Y; dX_3).$$

It is understood that the existence of all three integrals on the right of (60) is either supposed or proved as in the corresponding theorem.

6. Variationally orthogonal decompositions. It has been seen that if the vector $X(t)$ is decomposed into the sum of completely variationally orthogonal vectors then, in a large number of cases, (4) is correspondingly decomposed linearly. In case $X(t)$ is decomposed into the sum of two vectors which are variationally orthogonal but not necessarily completely so, the behavior of (4) is described by the following theorems.

THEOREM IX. *Suppose that* $\int_a^b F(Y; dX) = \int_a^b F(Y; dx_1, \cdots, dx_n)$ *and* $\int_a^b F(Y; 0, dx_2, \cdots, dx_n)$ *exist absolutely. Suppose that* F *satisfies the following condition: For every* $\epsilon > 0$, *there exists a* $\delta_\epsilon > 0$, *such that*

$$(61)\quad \left| \sum_{j=1}^{\infty} \{F(Y(\xi_j); z_{j1}, z_{j2}, \cdots, z_{jn}) - F(Y(\xi_j); w_{j1}, z_{j2}, \cdots, z_{jn})\} \right| < \epsilon \text{ if } \sum_{j=1}^{\infty} | z_{j1} - w_{j1} | < \delta_\epsilon.$$

Let $x_1(t) = x_{11}(t) + x_{21}(t)$, *where* x_{11}, x_{21} *are variationally orthogonal. Then*

$$(62)\quad \int_a^b F(Y; dX) = \int_a^b F(Y; dx_{11}, dx_2, \cdots, dx_n) + \int_a^b F(Y; dx_{21}, dx_2, \cdots, dx_n) - \int_a^b F(Y; 0, dx_2, \cdots, dx_n),$$

the existence of the first two integrals on the right being proved.

Theorem IX may be proved by an argument which is a modification of the proof of Theorem V and which will not be given here. It is found that the modification (61) of the condition (E) is all that is required in this case. A similar modification of the conditions (B), (C), (D), (D′), and (E′), in such a way that they apply only to the particular variable under consideration yield theorems which are analogous to Theorems VI, VII, VIII in the same way as Theorem IX is analogous to Theorem V.

Repeated applications of Theorem IX to successive variables yield

THEOREM X. *Suppose that* (4) *exists absolutely. Suppose that F satisfies condition* (*E*). *Let* $\mathbf{X}(t) = \mathbf{X}_1(t) + \mathbf{X}_2(t) + \mathbf{X}_3(t)$, *where* $\mathbf{X}_1$ *and* $\mathbf{X}_2$ *are variationally orthogonal vectors and* $\mathbf{X}_3 \equiv 0$. *Then*

$$(63) \qquad \int_a^b F(\mathbf{Y}; d\mathbf{X}) = \sum_{i_1=1}^{3} \cdots \sum_{i_n=1}^{3} (-1)^{k(i_1,\cdots,i_n)} \int_a^b F(\mathbf{Y}; dx_{i_1 1}, \cdots, dx_{i_n n}),$$

(*where* $k(i_1, \cdots, i_n) = \sum_{j=1}^{n} [i_j/3] =$ *number of values j for which* $i_j = 3$) *provided all the integrals occurring on the right of* (63) *exist absolutely.*

7. The Hellinger integral.[4]

The Hellinger integral is the special case of (4) for which $F = z_1{}^2/|z_2|$ so that the integral has the form

$$(64) \qquad \Phi(\mathbf{X}) = H_1(x_1, x_2) = \int_a^b (dx_1)^2/|dx_2|.$$

In dealing with these integrals it is always assumed that $x_2(t)$ is monotone non-decreasing in the interval $[a, b]$, so that the absolute value sign in the denominator of (64) is unnecessary, but it has been included here in order to make F satisfy condition (B). It is assumed that every interval of constancy of $x_2(t)$ is an interval of constancy of $x_1(t)$. Finally, the undetermined fraction 0/0 is defined to be 0.

The following known properties of (64) will be needed.

(i)[5] In order that (64) exist, it is necessary and sufficient that there exist a function $h(t)$, monotone non-decreasing in $[a, b]$, such that, if $a \leqq t < t' \leqq b$, then

$$(65) \qquad [x_1(t') - x_1(t)]^2 \leqq [x_2(t') - x_2(t)][h(t') - h(t)],$$

or, symbolically,

$$(65) \qquad (\Delta x_1)^2 \leqq (\Delta x_2)(\Delta h).$$

(ii)[6] Let $x_1(t) = g_1(t) - g_2(t)$ represent the decomposition (38), (39), (40) of $x_1(t)$ into the difference of two non-decreasing, variationally orthogonal functions. Then, if (64) exists, $H_1(g_1, x_2)$ and $H_1(g_2, x_2)$ exist and

$$H_1(x_1, x_2) = H_1(g_1, x_2) + H_1(g_2, x_2).$$

---

[4] Cf. Hellinger [4], Hahn [2], and Svenson [8].

[5] Cf. Hellinger [4], p. 26 f. or Svenson [8], p. 4.

[6] Svenson [8], Theorem IV, p. 29. This statement is also a consequence of one of the modifications of Theorem IX above and, in fact, holds for any decomposition of $x_1(t)$ into variationally orthogonal parts.

(iii)[7] If $x_1(t)$ is non-decreasing and if (64) exists, then the function $h(t)$ in (*i*) may be chosen so that $x_1(t)$ is a basis[8] of $h(t)$.

Now it will be shown that Theorems I, II, III, IV, and VIII are applicable to the Hellinger integral by proving

LEMMA 2. *The Hellinger integral* (64) *satisfies conditions* (*B*), (*C*) *and* (*D′*).

Proof. Since conditions (B) and (C) are obviously satisfied by $F = (z_1)^2/\,|\,z_2\,|$ it is only required to prove that (D′) also is satisfied.

Let $X(t) = \{x_1(t), x_2(t)\}$ be a vector function for which (64) exists. In view of (ii) it is sufficient to treat the case that $x_1(t)$, as well as $x_2(t)$, is non-decreasing. In this case, (i) and (iii) show that there exists a non-decreasing function $h(t)$, of which $x_1(t)$ is a basis, satisfying (65). Now, it is known[9] that if $x_1(t)$ is a basis of $h(t)$, then for every $\epsilon > 0$, there exists a $\delta_\epsilon > 0$ such that

$$\sum_{j=1}^{\infty} \Delta_j h < \epsilon \quad \text{whenever} \quad \sum_{j=1}^{\infty} \Delta_j x_1 < \delta_\epsilon.$$

Now suppose that $\sum_{j=1}^{\infty} [\Delta_j x_1 + \Delta_j x_2] < \delta_\epsilon$. Then,

$$\sum_{j=1}^{\infty} (\Delta_j x_1)^2/\Delta_j x_2 \leqq \sum_{j=1}^{\infty} \Delta_j h < \epsilon,$$

since $\Delta_j x_1 < \Delta_j x_1 + \Delta_j x_2$. But this is exactly condition (D′).

Thus it is seen, in particular, that the Hellinger integral is additive under completely variationally orthogonal decompositions $\mathbf{X} = \mathbf{X}_1 + \mathbf{X}_2$, whenever $\Phi(\mathbf{X}_1)$ and $\Phi(\mathbf{X}_2)$ exist. It will now be shown that this is always the case.

LEMMA 3. *Suppose that* (64) *exists where* $x_i(t)$, $i = 1, 2$, *are non-decreasing. Let* $x_i(t) = x_{1i}(t) + x_{2i}(t)$, *where* $x_{ij}(t)$, $i, j = 1, 2$, *are non-decreasing and where the vectors* $\mathbf{X}_1 = \{x_{11}, x_{12}\}$ *and* $\mathbf{X}_2 = \{x_{21}, x_{22}\}$ *are completely variationally orthogonal. Then* $H_1(x_{11}, x_{12})$ *and* $H_1(x_{21}, x_{22})$ *exist.*

Proof. Since $H_1(x_1, x_2)$ exists, by (i), there exists a monotone non-decreasing function $h(t)$ satisfying (65). By (65) and Schwarz's inequality

$$\Sigma\Delta_j x_1 \leqq \Sigma(\Delta_j x_2)^{\frac{1}{2}}(\Delta_j h)^{\frac{1}{2}} \leqq (\Sigma\Delta_j x_2)^{\frac{1}{2}}(\Sigma\Delta_j h)^{\frac{1}{2}},$$

so that, if S is any Borel set,

---

[7] Svenson [8], pp. 11-12.

[8] A monotone function $b(t)$ is called a basis of the function $f(t)$ if $V\{S; f\} = 0$ whenever $V\{S; b\} = 0$. Thus, $h(t)$ is clearly a basis of $x_1(t)$ by (65). Cf. Radon [6], p. 1318.

[9] Cf. Radon [6], p. 1319.

$$[V\{S;x_1\}]^2 \leqq V\{S;x_2\}V\{S;h\}. \tag{66}$$

Now by the definition (43) of complete variational orthogonality, there exists a set T such that

$$V\{T;x_{1i}\} = V\{[a,b)\,;\,x_{1i}\} \quad \text{and} \quad V\{T;x_{2i}\} = 0.$$

Thus, if $[t,t')$ is any half-open interval in $[a,b)$,

$$V\{[t,t')\,;\,x_{1i}\} = V\{[t,t')T;x_{1i}\} \tag{67}$$

and

$$V\{[t,t')T;x_i\} = V\{[t,t')T;x_{1i}\}. \tag{68}$$

In particular, (67) and (68) give

$$V\{[t,t')\,;\,x_{11}\} = V\{[t,t')T;x_1\},$$

or, using (66),

$$[V\{[t,t')\,;x_{11}\}]^2 \leqq V\{[t,t')T;x_2\}V\{[t,t')T;h\}. \tag{69}$$

But, by (68),

$$V\{[t,t')T;x_2\} = V\{[t,t')T;x_{12}\}. \tag{70}$$

Together, (69) and (70) imply

$$[V\{[t,t')\,;x_{11}\}]^2 \leqq V\{[t,t')\,;\,x_{12}\}V\{[t,t')\,;\,h\}$$

or

$$(\Delta x_{11})^2 \leqq (\Delta x_{12})(\Delta h).$$

Hence, by (i), $H(x_{11},x_{12})$ exists. The existence of $H(x_{21},x_{22})$ is shown similarly.

Some of the results of this section are summarized, for the particular case of the Lebesgue decomposition, in

COROLLARY 2. *Suppose that* (64) *exists. Let*

$$x_i(t) = a_i(t) + s_i(t) + p_i(t), \qquad (i = 1,2),$$

where a_i is absolutely continuous, s_i is purely singular, and p_i is purely discontinuous. Then

$$H_1(x_1,x_2) = H_1(a_1,a_2) + H_1(s_1,s_2) + H_1(p_1,p_2)$$

or

$$\int_a^b (dx_1)^2/dx_2 = \int_a^b [(x'_1(t))^2/x'_2(t)]dt + \int_a^b (ds_1)^2/ds_2$$
$$+ \Sigma[x_1(\rho_i+0) - x_1(\rho_i-0)]^2/[x_2(\rho_i+0) - x_2(\rho_i-0)],$$

where the last sum is taken over all discontinuity points ρ_i of $x_1(t)$.

It may be mentioned that the results of this section also apply to the generalisation,[10] $\int_a^b |dx_1|^p / |dx_2|^{p-1}$, $p > 1$, of Hellinger's integral.

8. The Hilbert integral.[11] The Hilbert integral is the special case of (1) for which $F = |z_1 z_2|^{\frac{1}{2}}$, so that the integral has the form

$$\Phi(X) = H_2(x_1, x_2) = \int_a^b |dx_1 dx_2|^{\frac{1}{2}}. \tag{71}$$

This integral exists whenever $X(t) = \{x_1(t), x_2(t)\}$ is of bounded variation. It will now be shown that Theorems I, II, III, IV and VI are applicable to this integral by proving

LEMMA 4. *The Hilbert integral* (71) *satisfies condition* (E').

Proof. Let $X_i(t) = \{x_{i1}(t), x_{i2}(t)\}$, $i = 1, 2$, be vectors of bounded variation, so that $H_2(x_{11}, x_{12})$ and $H(x_{21}, x_{22})$ exist. Let $M = \max\limits_{i,j=1,2} V\{[a, b]; x_{ij}\}$. Now, from

$$|a + b|^{\frac{1}{2}} |c + d|^{\frac{1}{2}} \leqq |ac|^{\frac{1}{2}} + |ad|^{\frac{1}{2}} + |bc|^{\frac{1}{2}} + |bd|^{\frac{1}{2}},$$

one obtains

$$\begin{aligned} |\Delta x_{11}\Delta x_{12}|^{\frac{1}{2}} - |\Delta x_{21}\Delta x_{22}|^{\frac{1}{2}} \leqq{} & |\Delta x_{12} - \Delta x_{22}|^{\frac{1}{2}} |\Delta x_{21}|^{\frac{1}{2}} \\ & + |\Delta x_{11} - \Delta x_{21}|^{\frac{1}{2}} |\Delta x_{22}|^{\frac{1}{2}} + |\Delta x_{12} - \Delta x_{22}|^{\frac{1}{2}} |\Delta x_{11} - \Delta x_{21}|^{\frac{1}{2}}, \end{aligned} \tag{72}$$

by letting $\Delta x_{11} = a + b$, $\Delta x_{12} = c + d$, $\Delta x_{21} = a$ and $\Delta x_{22} = c$. Using (72) and the Schwarz inequality, one has

$$\sum_{j=1}^{\infty} \{|\Delta_j x_{11}\Delta_j x_{12}|^{\frac{1}{2}} - |\Delta_j x_{21}\Delta_j x_{22}|^{\frac{1}{2}}\} \leqq 4M^{\frac{1}{2}} \{\sum_{j=1}^{\infty}\sum_{k=1}^{2} |\Delta_j x_{1k} - \Delta_j x_{2k}|\}^{\frac{1}{2}} \tag{73}$$

where $\Delta_j x_{ik} = x_{ik}(t'_j) - x_{ik}(t_j)$ and $\{[t_j, t'_j)\}$ is any set of non-overlapping half-open intervals on $[a, b]$.

By repeating the above argument with the subscripts 1 and 2 interchanged, it is seen that (73) still holds if the absolute value of the first sum is taken. Thus

$$\left|\sum_{j=1}^{\infty} \{|\Delta_j x_{11}\Delta_j x_{12}|^{\frac{1}{2}} - |\Delta_j x_{21}\Delta_j x_{22}|^{\frac{1}{2}}\}\right| < \epsilon \text{ if } \sum_{j=1}^{\infty}\sum_{k=1}^{2} |\Delta_j x_{1k} - \Delta_j x_{2k}| < \delta_\epsilon, \tag{74}$$

where $\delta_\epsilon = \epsilon^2/16M$. But (74) is precisely condition (E′).

Recalling the fact that condition (E′) implies condition (D′) it is seen that the theorems mentioned above are applicable to the Hilbert integral (71). Again restating the particular case of the Lebesgue decomposition, one has

---

[10] F. Riesz [6].

[11] Hellinger [4], p. 31 f.

COROLLARY 3. *Let $x_i(t) = a_i(t) + s_i(t) + p_i(t)$, $i = 1, 2$, where a_i is absolutely continuous, s_i is purely singular, and p_i is purely discontinuous. Then*

$$H_2(x_1, x_2) = H_2(a_1, a_2) + H_2(s_1, s_2) + H_2(p_1, p_2)$$

or

$$\int_a^b |\, dx_1 dx_2 \,|^{\frac{1}{2}} = \int_a^b |\, x'_1(t) x'_2(t) \,|^{\frac{1}{2}} dt + \int_a^b |\, ds_1 ds_2 \,|^{\frac{1}{2}} \\ + \Sigma \,|\, x_1(\rho_i + 0) - x_1(\rho_i - 0) \,|^{\frac{1}{2}} \,|\, x_2(\rho_i + 0) - x_2(\rho_i - 0) \,|^{\frac{1}{2}},$$

where the last sum is taken over all common discontinuity points ρ_i of $x_1(t)$ and $x_2(t)$.

9. Riemannian arc length. The length of curves in Riemannian spaces is given by the integral

$$(75) \qquad \int_a^b \left\{ \sum_{i,k=1}^{n} g_{ik}(x_1, \cdots, x_n)\, dx_i dx_k \right\}^{\frac{1}{2}},$$

where $g_{ik}(t) = g_{ik}(x_1(t), \cdots, x_n(t))$ is the matrix of a positive definite quadratic form. It is assumed that $g_{ik}(t)$ is continuous on $[a, b]$. Then the integral (75) exists whenever $\mathbf{X}(t) = \{x_1(t), \cdots, x_n(t)\}$ is of bounded variation. It will be shown that Theorems I, II, III, IV, VII, IX and X are applicable to these integrals (75) by proving

LEMMA 5. *The integral (75) satisfies conditions (B), (C) and (D).*

Proof. First, the condition (B) is obviously satisfied.

To see that condition (D) is satisfied, let $M = \max |\, g_{ik}(t) \,|$ for $a \leqq t \leqq b$, $i, k = 1, \cdots, n$. Then

$$\left\{ \sum_{i,k=1}^{n} g_{ik}(t) z_i z_k \right\}^{\frac{1}{2}} \leqq M^{\frac{1}{2}} \sum_{i=1}^{n} |\, z_i \,|.$$

Finally, the inequality required in (C) is known [12] to hold in this case.

COROLLARY 4. *Let $x_i(t) = a_i(t) + s_i(t) + p_i(t)$, $i = 1, \cdots, n$, where a_i is absolutely continuous, s_i is purely singular and p_i is purely discontinuous. Then*

$$(76) \qquad \int_a^b \left\{ \sum_{i,k=1}^{n} g_{ik}(\mathbf{X})\, dx_i dx_k \right\}^{\frac{1}{2}} = \int_a^b \left\{ \sum_{i,k=1}^{n} g_{ik}(\mathbf{X})\, da_i da_k \right\}^{\frac{1}{2}} \\ + \int_a^b \left\{ \sum_{i,k=1}^{n} g_{ik}(\mathbf{X})\, ds_i ds_k \right\}^{\frac{1}{2}} + \int_a^b \left\{ \sum_{i,k=1}^{n} g_{ik}(\mathbf{X})\, dp_i dp_k \right\}^{\frac{1}{2}}.$$

---

[12] This inequality is equivalent to inequality 29 in Hardy, Littlewood and Pólya [3], p. 33.

10. Euclidean arc length. The ordinary Euclidean arc length is, of course, the special case of the above for which $g_{ik} = \delta_{ik}$. However, the results obtained take such a simple form, in this case, that it seems worthwhile to enumerate them. For simplicity of statement, these results will be formulated in the case $n = 2$.

COROLLARY 5. *Let* $x_i(t) = a_i(t) + s_i(t) + p_i(t)$, $i = 1, 2$, *where* a_i *is absolutely continuous,* s_i *is purely singular, and* p_i *is purely discontinuous. Then*

$$(77) \quad \int_a^b \{(dx_1)^2 + (dx_2)^2\}^{\frac{1}{2}} = \int_a^b \{(x'_1(t))^2 + (x'_2(t))^2\}^{\frac{1}{2}} dt$$

$$+ \int_a^b \{(ds_1)^2 + (ds_2)^2\}^{\frac{1}{2}} + \Sigma\{[x_1(\rho_i + 0) - x_i(\rho_i - 0)]^2$$

$$+ [x_2(\rho_i + 0) - x_2(\rho_i - 0)]^2\}^{\frac{1}{2}},$$

where the last sum is taken over all discontinuity points ρ_i *of* $x_1(t)$ *or of* $x_2(t)$.

COROLLARY 6. *Suppose* $\mathbf{S}(t) = \{s_1(t), s_2(t)\}$ *is a continuous vector function of bounded variation, such that, for every* t *on* $[a, b]$, *either* $s'_1(t) = 0$ *or* $s'_2(t) = 0$. *Then*

$$\int_a^b \{(ds_1)^2 + (ds_2)^2\}^{\frac{1}{2}} = V\{[a, b); s_1\} + V\{[a, b); s_2\}.$$

In fact, in the decomposition $\mathbf{S} = \mathbf{S}_1 + \mathbf{S}_2$, where $\mathbf{S}_1(t) = \{s_1(t), 0\}$ and $\mathbf{S}_2(t) = \{0, s_2(t)\}$, the two vectors $\mathbf{S}_1$ and $\mathbf{S}_2$ are completely variationally orthogonal.

COROLLARY 7. *The length of the curve* $y = y(x)$, *where* $y(x)$ *is a continuous function of bounded variation in* $[a, b]$, *is given by*

$$\int_a^b \{1 + (y'(t))^2\}^{\frac{1}{2}} dt + V\{[a, b); y\} - \int_a^b |y'(t)| \, dt.$$

This is a consequence of Corollary 5, for in this case $a_1(t) = t$, $a_2(t) = \int_a^t y'(u)\,du$, $s_1(t) = 0$, $s_2(t) = y(t) - a_2(t)$.

COROLLARY 8. *The length of the curve* $y = y(x)$, *where* $y(x)$ *is a purely singular monotone function in* $[a, b]$ *is given by*

$$|y(b) - y(a)| + |b - a|.$$

COROLLARY 9. *Let* $x_1(t) = g(t) - h(t)$ *be the decomposition* (38), (39), (40) *of* $x_1(t)$ *into the difference of two monotone non-decreasing, variationally orthogonal functions. Then*

$$\int_a^b \{(dx_1)^2 + (dx_2)^2\}^{\frac{1}{2}} = \int_a^b \{(dg)^2 + (dx_2)^2\}^{\frac{1}{2}} + \int_a^b \{(dh)^2 + (dx_2)^2\}^{\frac{1}{2}} - V\{[a, b)\,;\, x_2\}.$$

This corollary is a consequence of Theorem VIII.

COROLLARY 10. *Let $x_i(t) = g_i(t) - h_i(t)$ be the decomposition of $x_i(t)$ into the difference of two monotone non-decreasing variationally orthogonal functions. Then*

$$\begin{aligned}\int_a^b \{(dx_1)^2 + (dx_2)^2\}^{\frac{1}{2}} = \int_a^b &\{(dg_1)^2 + (dg_2)^2\}^{\frac{1}{2}} \\ &+ \int_a^b \{(dh_1)^2 + (dh_2)^2\}^{\frac{1}{2}} + \int_a^b \{(dg_1)^2 + (dh_2)^2\}^{\frac{1}{2}} \\ &+ \int_a^b \{(dh_1)^2 + (dg_2)^2\}^{\frac{1}{2}} \\ &- V\{[a, b)\,;\, x_1\} - V\{[a, b)\,;\, x_2\}.\end{aligned}$$

This corollary is a consequence of Theorem IX.

In Corollaries 9 and 10 it is clear that the variational orthogonality of g_i and h_i is all that is needed to insure the result.

THE JOHNS HOPKINS UNIVERSITY,
UNIVERSITY OF WISCONSIN.

BIBLIOGRAPHY.

[1] P. Finsler, "Über Kurven und Flächen in allgemeinen Räumen," Dissertation, Göttingen (1918).

[2] H. Hahn, "Über die Integrale des Herrn Hellinger und die Orthogonalinvarianten der quadratischen Formen von unendlich vielen Veränderlichen," *Monatshefte für Mathematik und Physik*, vol. 23 (1912), pp. 161-224.

[3] G. H. Hardy, J. E. Littlewood, and G. Pólya, *Inequalities*, Cambridge (1934).

[4] E. Hellinger, "Die Orthogonalinvarianten quadratischer Formen von unendlichvielen Variablen," Dissertation, Göttingen (1907).

[5] J. Radon, "Theorie und Anwendungen der absolut additiven Mengenfunctionen," *Sitzungsberichte der Kaiserlichen Akademie der Wissenschaften, Wien, Math.-Nat. Klasse*, vol. 122 (1913), Abth. 2a, pp. 1295-1438.

[6] F. Riesz, "Untersuchungen über Systeme integrierbarer Funktionen," *Mathematische Annalen*, vol. 69 (1910), pp. 449-497.

[7] S. Saks, *Théorie de l'intégrale*, Warsaw (1933).

[8] E. Svenson, "Beiträge zur Theorie gewisser Integraltypen," *Abhandlungen der Herder-Gesellschaft und des Herder-Instituts zu Riga*, vol. 5 (1936), no. 6.

ON TRANSLATIONS IN GENERAL PLANE GEOMETRIES.*

By HERBERT BUSEMANN.

In a well-known paper, Hilbert[1] has characterized the Euclidean and hyperbolic plane geometries by mere group and continuity axioms. He gets all the motions *at once* by requiring the existence of sufficiently many *rotations.* The present paper tries to point out how the existence of more and more *translations gradually* specializes the rather general metric it starts with to a Desarguesian geometry (Minkowskian or hyperbolic).

We require our initial space to be a " Geradenraum " in Menger's terminology.[2] The exact definition will be found in section 1. It is essentially a metric space with exactly one shortest line (s. l.) of infinite length passing through two given points. We therefore call a Geradenraum an SL space. In order to be able to formulate the results, we must have the concept of an asymptote. If g is any shortest line, P a point not on g, and Q a point traversing g in a certain direction $\overrightarrow{g}$ ($\overleftarrow{g}$), the s. l. connecting P and Q always tends to a limit s. l., which we call an *asymptote* to $\overrightarrow{g}$, preserving the word *parallel* for the case where the two asymptotes to $\overrightarrow{g}$ and $\overleftarrow{g}$ through P coincide. Section 1 gives those properties of these asymptotes and of limit circles which we shall need later on.

A two-dimensional SL space Σ will be called a *plane* since it is homeomorphic to the Euclidean plane.[3] As usually, we say the metric of Σ is *Desarguesian* if it is possible to map Σ topologically on the Euclidean plane or a convex part K of it in such manner that the shortest lines are transformed into straight lines or into the intersections of straight lines with K.[4] By a motion of Σ we mean any one-to-one mapping of Σ onto itself which preserves distance and, in particular, by a *translation along a shortest line* g we mean a motion which transforms each of the half planes of Σ defined by g into itself.

We first assume that to each pair A, B of points on a fixed s. l. g there

---

* Received April 19, 1937.

[1] Reprinted in [5], Anhang IV. The numbers [n] refer to the references on p. 256.

[2] [7] especially pp. 100-113.

[3] See [2].

[4] A proof for the fact that the validity of Desargues' Theorem is necessary and sufficient for the existence of such a mapping, can be found in [11] §§ 3, 4.

exists a translation along g carrying A into B; these translations form a group G. In section 2 we ascertain which of the usual properties of translations hold, and show by examples that others do not. The main result is that *the images of a fixed point A under the translations of G form a curve which together with g bounds a convex domain.* It follows then, for instance, that a parallel h to g is equidistant from g and that the translations along g can also be regarded as translations along h; but such a geometry is not necessarily Desarguesian, even if the parallel axiom holds throughout the plane. On the other hand, it can be Desarguessian with the parallel axiom holding only with respect to g.

The availability of G does not imply the existence of translations along a non-parallel asymptote to $\overleftarrow{g}$ (or $\overrightarrow{g}$). By assuming their presence one therefore gets a much more specialized metric (section 3), in which translations along each asymptote a' to $\overleftarrow{g}$ exist and *the hyperbolic formula for the arc of the limit circle to the direction $\overleftarrow{g}$ between a' and g holds.* Nevertheless an example will show that the geometry is *not necessarily Desarguesian.*

However, *as soon as one requires the existence of translations along two s. l. where the one is neither an asymptote nor parallel to the other, we get a Desarguesian metric.* We find, namely, (section 4) that *the metric is either Minkowskian*[5] *or hyperbolic.* In both cases there exist translations along each straight line.

1. Asymptotes and limit spheres in SL spaces. The exact definition of an SL space is as follows: A complete metric space with the distance function $r(x,y)$ is an SL space if it satisfies the conditions:

I. To each pair of points A, B there exists exactly one point C, the center of A and B for which $r(A,C)+r(C,B)=r(A,B)$, $r(A,C)=\dfrac{r(A,B)}{2}$.

II. To each pair of distinct points A, B there exist exactly two points D and D' such that B is the center of A and D and A the center of B and D'.

All points X satisfying the relation

$$r(A,X)+r(X,B)=r(A,B)$$

form a point set (designated by $\overline{AB}$ or $\overline{BA}$), which is homeomorphic to an Euclidean straight line segment.[6] If $X \subset \overline{AB}$ and $X \neq A$, $X \neq B$ we say that X is between A and B. The set of points Y (resp. Z) for which B is between A and Y (resp. A between B and Z) together with $\overline{AB}$ is homeomorphic to a Euclidean ray and will be designated by $\underrightarrow{AB}$, resp. $\underrightarrow{BA}$. We put

---

[5] The original definition is in [8], chapter I, compare also p. 234 of this paper.

[6] *loc. cit.* [2].

$$\underline{AB} = \underrightarrow{AB} + \underrightarrow{BA}$$

and call $\underline{AB}$ a *shortest line* (*s. l.*). $\underline{AB}$ is homeomorphic to a Euclidean straight line.[6] Then the following theorem holds:

(1. 1) *Through two different points passes exactly one* s. l. *Hence two different* s. l. *intersect in at most one point.*

As usually we put for any two point sets

$$r(\alpha, \beta) = \underset{X \subset \alpha,\; Y \subset \beta}{\text{G. l. b.}}\, r(X, Y)$$

We say that the point sets α_ν converge to the set α if α contains all limit points of sequences $\{P_\nu\}$ with $P_\nu \subset \alpha_\nu$ and if to any point O, any positive number r, and any $\epsilon > 0$, a number $N(\epsilon, O, r)$ can be found such that for $\nu > N$ and each point $X \subset \alpha$ with $r(X, O) < r$ the inequality

$$r(X, \alpha_\nu) < \epsilon$$

holds. Then we have

(1. 2) *If* $A_\nu \to A$ *and* $B_\nu \to B$, *then* $\overline{A_\nu B_\nu} \to \overline{AB}$; *and if* $A \neq B$, *also*

$$\underrightarrow{A_\nu B_\nu} \to \underrightarrow{AB}, \qquad \underrightarrow{B_\nu A_\nu} \to \underrightarrow{BA}, \qquad \underline{A_\nu B_\nu} \to \underline{AB}.$$

If, for a given point P, a point $Q < \alpha$ exists with

$$r(P, Q) = r(P, \alpha)$$

we call Q a *foot* of P on α. Then the following lemma is an immediate consequence of the triangle inequality.

(1. 3) *If* $R_\nu \to R$, $\alpha_\nu \to \alpha$, *and* T_ν *is a foot of* R_ν *on* α_ν, *then each accumulation point* T *of* T_ν *is a foot of* R *on* α.

Let g be any s. l., $P \not\subset g$, $A \subset g$, and X a variable point on g. $r(X, A) \to \infty$ implies $r(P, X) \to \infty$; hence, P has at least one foot Q on g. A point R between P and Q has Q as its only foot on g; for, from

$$r(R, Q') \leq r(R, Q), \qquad Q' \subset g$$

would follow

$$r(P, Q') < r(P, R) + r(R, Q) = r(P, g).$$

We shall have to use later the fact:

(1.4) *In an* SL *space of dimension greater than* 1, *there exist to each s.l. g points whose distance from g is arbitrarily great.*

To prove this we show that a sphere with radius a whose center C is on g[7] contains points P with $r(P, g) \geq a/2$. Let B_1, B_2 be the points on g with $r(B_1, C) = r(B_2, C) = a$, B'_1 the center of $\overline{B_1C}$, B'_2 the center of $\overline{B_2C}$ and A an arbitrary point not on g. We draw $\overline{B_1A}$ and $\overline{B_2A}$ and consider the set σ formed by all rays $\overrightarrow{CX}$ with $X \subset \overline{B_1A} + \overline{AB_2}$. Let K be the intersection of the sphere with σ; K is homeomorphic to a Euclidean semicircle. The feet on g of a point $P \subset K$ near B_i are near B_i $(i = 1, 2)$. If one had $r(P, g) < a/2$ for all $P \subset K$ no foot on g of a point $P \subset K$ could belong to $\overline{B'_1B'_2}$. It is easy to see that there must be a point P_0 with two feet Q_1, Q_2 on g such that $Q_1 \subset \overrightarrow{B'_1B_1}$, $Q_2 \subset \overrightarrow{B'_2B_2}$ and one would have

$$a > r(P_0, Q_1) + r(P_0, Q_2) \geq r(Q_1, Q_2) \geq a.$$

(1.4), together with the preceding statement, gives:

(1.5) *In an* SL *space of dimension greater than* 1, *to a given s.l. g and a given* $N > 0$ *points P can be found with a unique foot on g and* $r(P, g) = N$.

We now consider an oriented s.l. g and write $Q < Q'$ for two points Q and Q' on g if Q' follows Q. If $P < Q < Q'$ it follows from the triangle inequality that

$$K(Q, P) \cdot K(Q', P) = P \tag{1.6}$$

($\alpha \cdot \beta$ means the product of the point sets α and β) where $K(X, Z)$ means the sphere with the center X and through Z, i.e. with the radius $r(X, Z)$. More exactly, $K(P, Q) - P$ lies in the interior of $K(P, Q')$. From this we conclude: if $P < Q_1 < Q_2 < \cdots$ and $r(P, Q_\nu) \to \infty$, the spheres $K(Q, P)$ tend to a limit set $L(P, g)$, which does not depend on the choice of the sequence $\{Q_\nu\}$.[8] We call $L(P, g)$ the *limit sphere through P with or to the center ray g*. In terms of the metric of the space, $L(P, g)$ can be characterized as follows: it consists of those points R for which

---

[7] By the sphere with radius a and center C we mean the set of points Z with $r(Z, C) = a$.

[8] For detailed proofs of this and the following statements see [3] § 2.

$$(1.7) \qquad r(Q_\nu, P) - r(Q_\nu, R) \to 0$$

for each sequence $Q_1 < Q_2 < \cdots$ with $r(P, Q) \to \infty$.

One proves with the help of (1.7) that the limit spheres with the center ray g cover the space. $L(P, g)$ decomposes the space into two parts, the interior and the exterior of $L(P, g)$, the former consisting of those points which, for $P < Q$ and sufficiently large $r(P, Q)$, are in the interior of $K(Q, P)$.

We call the point set μ equidistant from the point set ν if, for any two points P_1, P_2 of μ one has

$$r(P_1, \nu) = r(P_2, \nu) = r(\mu, \nu).$$

If μ is equidistant from ν, then ν is not necessarily equidistant from ν. But for limit spheres one has

(1.8) *Limit spheres with the same center ray g are equidistant from each other.* Hence exactly one limit sphere with g as center ray passes through a given point.

With the help of these results we can prove a lemma which will be applied frequently in the sequel.

(1.9) LEMMA. *If the s. l. s and a intersect in T then the parts of a and s outside any sphere around T with positive radius have positive distance.*

Let us assume, on the contrary, that there exists a sequence $\{R^*_\nu\}$ on a and a sequence $\{S^*_\nu\}$ on s, such that

$$\begin{aligned} r(R^*_\nu, S^*_\nu) \to 0 \quad \text{but} \quad & r(R^*_\nu, T) > \delta > 0 \\ & r(S^*_\nu, T) > \delta > 0. \end{aligned}$$

Then we must have $r(R^*_\nu, T) \to \infty$.

Let A be a point on a different from T. Either on $\overrightarrow{AT}$ or on $\overrightarrow{TA}$ are infinitely many points R_ν of $\{R^*_\nu\}$; suppose they are on $\overrightarrow{TA}$, and call $\{S_\nu\}$ the subsequence of $\{S^*_\nu\}$ corresponding to $\{R_\nu\}$. The two limit spheres $L(A, \overrightarrow{TA})$ and $L(A, \overrightarrow{AT})$ have no common interior point. Since $r(S_\nu, R_\nu) \to 0$, the s. l. s intersects both these limit spheres, the former at B, say, and the latter at B', with $B' \subset \overline{TB}$. On account of (1.6) the sphere $K(T, A)$ intersects the segment $\overline{TB'}$ in a point B'' between T and B'; hence, $r(T, B) > r(T, A)$. On the other hand, we have $| r(B, S_\nu) - r(B, R_\nu) | < r(R_\nu, S_\nu)$ and

$$r(B, R_\nu) - r(A, R_\nu) \to 0 \qquad (\text{see } (1.7)).$$

Hence

$$r(B, S_\nu) - [r(A, R_\nu) + r(R_\nu, S_\nu)] \to 0$$

and

$$r(T, S_\nu) - [r(T, R_\nu) + r(R_\nu, S_\nu)] \to r(T, B) \; - r(T, A) > 0,$$

which for large ν contradicts the triangle inequality.

From reference [3], section 2, we take the following theorem:

(1.10) *An arbitrary point A has a unique foot on each limit sphere. The feet of A on the different limit spheres with the same center ray g form a s. l. a.*

We call a the *asymptote* to g through A. Two asymptotes to g are either identical or disjoined. In order to justify the use of this term we show

(1.11) *If $\{Q^*_\nu\} < g$, $Q^*_1 < Q^*_2, \cdots$, and $r(Q^*_1, Q^*_\nu) \to \infty$, then $\underline{AQ^*_\nu}$ tends to the asymptote a to g through A.*

Proof. Let $L(P, g)$ be the limit sphere with the center ray g through A, and $\{Q_\nu\}$ a subsequence of $\{Q^*_\nu\}$ for which $\underline{AQ_\nu}$ converges to a s. l. b. (That $\{Q_\nu\}$ exists follows from (1.2)). We have to prove that b coincides with a. For $r(A, Q_\nu) > 1$ we can choose the points R_ν and T_ν on $\overline{AQ_\nu}$ such that $r(A, R_\nu) = 1$ and $r(Q_\nu, P) = r(Q_\nu, T_\nu)$ (that $r(Q_\nu, A) > r(Q_\nu, P)$ follows from (1.1)). The points R_ν tend to a point R on b with $r(A, R) = 1$, R is not on $L(P, g)$. T_ν is the foot of R on $K(Q_\nu, P)$. We therefore conclude from (1.3) and $K(Q_\nu, P) \to L(P, \vec{g})$ that each limit point of $\{T_\nu\}$ must be a foot of R on $L(P, g)$. (1.10) shows that R has only one foot T on $L(P, g)$, namely the point where the asymptote $\underline{RT}$ to g through R intersects $L(P, g)$. Therefore,

$$b = \lim \underline{QA} = \lim \underline{RT} = \underline{RT}$$

Hence b is an asymptote to g and from $b \cdot a \supset A$ follows $b \equiv a$.

We can formulate our results thus:

(1.12) *If Q traverses the s. l. g in a certain direction $\vec{g}$ and $P \not\subset g$, then $\underline{PQ}$ tends to a limit s. l. a, the asymptote to $\vec{g}$ through P. The asymptote to $\vec{g}$ through a point of* a *coincides with* a. *A converging sequence of asymptotes to $\vec{g}$ tends to an asymptote to $\vec{g}$.*

From now on we suppose that our SL space, indicated by Σ, has dimension 2. It is therefore hemeomorphic to the Euclidean plane.[3] The spheres are homeomorphic to Euclidean circles; we therefore speak of *circles* and *limit circles* instead of spheres and limit spheres. Through each point A not on a s. l. g we have two (possibly coincident) asymptotes to g. They determine two angles, an open one and closed one. The former is determined by the property that it contains all the s. l. through A intersecting g and no others; the latter is the complement to the former. If the two asymptotes coincide, the closed angle consists of exactly one s. l., which we call a *parallel* to g.

Let $\vec{a}$ be an asymptote to $\vec{g}$. The question arises as to whether the family of the limit circles to $\vec{a}$ is (as in the hyperbolic and Euclidean geometries) always identical with that of the limit circles to $\vec{g}$. We shall see that only part of this is true:

(1.13) *Let π^* be that one of the two half-planes defined by a which is contained in one of the half-planes determined by g, π^{**} the other one. Then.* for each $P \subset \vec{g}$.

$$L(P,\vec{g})\pi^* \equiv L(T,\vec{a})\pi^* \qquad (T = L(P,\vec{g}) \cdot a),$$

and $L(T,\vec{a})\pi^{**}$ is in the interior or on $L(P,\vec{g})$.

To prove this, we first state that for each pair of distinct points A, B and each pair of positive numbers a, b, with $a + b > r(A,B)$, exactly two points T and T', one on each side of $\underline{AB}$, exist such that

$$(1.14) \qquad r(A,T) = r(A,T') = a, \qquad r(B,T) = r(B,T') = b.$$[9]

If $b = a$, we can say, furthermore, that $\overline{TT'}$ lies in the interior of the quadrangle $ATBT'$. For, otherwise, we would have

$$r(T,A) = r(T,B), \qquad r(T',A) = r(T',B)$$

with A and B on the same side of $\underline{TT'}$.

We now consider any point P of g. $L(P,\vec{g})$ may intersect a in T. Let A be any point following P on g, and B the point on a with $r(B,T) = r(A,P)$. Except for the point T, the circle $K(B,T)$ lies in the interior of $L(P,\vec{g})$, since T is the unique foot of B on $L(P,\vec{g})$. This statement contains the second part of (1.13). On account of (1.14) the circles $K(A,P)$ and $K(B,T)$ intersect exactly twice for sufficiently large $r(A,P)$. Since $K(A,P)$

---

[9] Compare [2].

is (except for P) in the interior of $L(P,\vec{g})$, one of the intersections, for large $r(A,P)$, must be between a and g near the arc $\overset{\frown}{PT}$ of $L(P,\vec{g})$, and the other outside of an arbitrarily large circle around T. Hence, for large $r(A,P)$, an arbitrary great portion of $K(B,T)\cdot\pi^*$ lies between $K(A,P)$ and $L(P,\vec{g})$, which proves the remainder of the assertion.

It follows that:

(1.15) *The asymptotes to $\vec{a}$ in π^* are also asymptotes to $\vec{g}$.*
This can, of course, be easily seen without using (1.13).

We now show by an example that, even in a Desarguesian geometry with the Euclidean parallel axiom it can happen that in the above notation $L(T,\vec{a})\pi^{**}$ lies wholly in the interior of $L(P,\vec{g})$.

We recall the definition of a Minkowskian geometry: Let $r=\lambda(\phi)$ be a convex curve in the strict sense in the Cartesian (r,ϕ) or (x,y)-plane with $r=0$ as center. We define the distance between two points $P_1=(x_1,y_1)$, $P_2=(x_2,y_2)$ to be the number

$$\bar{r}(P_1,P_2)=\sqrt{(x_1-x_2)^2+(y_1-y_2)^2}\,\frac{1}{\lambda(\phi)}$$

where $+\phi$ or $-\phi$ is the direction of the straight line $\underline{P_1P_2}$, i.e.

$$\phi=\pm\operatorname{arc}tg\,\frac{y_2-y_1}{x_2-x_1}.$$

As shown by Minkowski[5] $\bar{r}(P_1,P_2)$ satisfies all our conditions and the whole Euclidean straight lines are the s.l. of our space. Now let $r=\lambda(\phi)$ be the curve

$$x^2+\left(y+\frac{1}{\sqrt{2}}\right)^2=1 \quad \text{for } y\geq 0,$$
$$x^2+\left(y-\frac{1}{\sqrt{2}}\right)^2=1 \quad \text{for } y\leq 0.$$

The limit circles to center rays with directions different from 0 and π are straight lines. The limit circle to the positive x-axis through 0 is the curve

$$(*)\qquad \begin{cases} y=x & \text{for } \ y\geq 0,\\ y=-x & \text{for } \ y\leq 0.\end{cases}$$

The limit circle to a center ray $y=\beta$ parallel to the positive x-axis through the point (α,β) is obtained from $(*)$ by a translation which carries $(0,0)$ into (α,β). One sees: if $\beta>0$ and $\alpha=\beta$ (for $\beta<0$, $\alpha=-\beta$) the two limit circles are identical only for $y \geq \beta$ $(y \leq \beta)$ and the second is in the interior of (*) for $y < \beta$ $(y > \beta)$.

I do not know if the concept of an asymptote is always symmetric, i. e. if from the fact that a is an asymptote to g always follows that g is an asymptote to a. A sufficient condition is that the limit spheres to a are also limit spheres to g, but this condition is not necessary, as we have just seen. Another sufficient condition is given by

(1.16) *If the distance between two non-intersecting s. l.'s a and b vanishes, then a is an asymptote to b and b to a.*

For, there exists a sequence of points R_ν on b, tending to infinity in a certain direction, say $\vec{b}$, and a sequence of points S on a with $r(S_\nu, R_\nu) \to 0$. We select an arbitrary point A on a. The s. l. $\underline{AR_\nu}$ tend to the asymptote c to b through A. If c is not identical with a, it must intersect $\overline{S_\nu R_\nu}$ in a point S'_ν and we would have $r(S_\nu, S'_\nu) \to 0$ in contradiction to (1.9).

The converse to (1.16) cannot always hold, since it is not true in Euclidean geometry. But one could conjecture that it is true if the parallel axiom of hyperbolic geometry holds throughout the plane, especially in a Desarguesian geometry in a bounded part of the Euclidean plane. To show that this *is not so*, we recall the definition of a certain geometry introduced by Hilbert.[10] Let K be any bounded, closed, convex curve in the plane. The distance between two points A, B interior to K is defined as follows: Let $\overrightarrow{AB}$ intersect K in Y and $\overrightarrow{BA}$ in X. Designating the Euclidean distance by $e(\ )$ we put

$$r^*(A, B) = \log \frac{e(A, Y) \cdot e(B, X)}{e(B, Y) \cdot e(A, X)}.$$

Hilbert proves that $r^*(A, B)$ satisfies our conditions if K is convex in the strict sense and that otherwise the straight lines are shortest lines but not necessarily the only s. l. Therefore

$$r^*(A, B) + e(A, B)$$

is, for each bounded, convex curve K, a distance function which defines an SL space. Choosing K as a triangle, one sees that the s. l. issuing from a vertex of K are asymptotes to each other, but the distance between any two distinct ones among these s. l. is positive. One should note that the distance of the "other ends" of two such s. l. is finite.

---

[10] See [5], Anhang I.

2. Geometries with a group of translations along one shortest line. Let g be a s. l. in the two dimensional SL space Σ, π_1 and π_2 the two half-planes into which g decomposes Σ. A one-to-one transformation of Σ into itself, which preserves the distances of corresponding pairs of points and transforms π_1, π_2 into themselves will be called a *translation of* Σ *along* g. Such a translation transforms a s. l. into a s. l.

To a given pair of points A, B on g there exists at most one translation along g carrying A into B. For let C be any point on g different from A, then any translation γ along g carrying A into B transforms C into the point $D \subset g$ with $\gamma(C, D) = \gamma(A, B)$ and either

$$\overrightarrow{AB} \supset \overrightarrow{CD} \quad \text{or} \quad \overrightarrow{CD} \supset \overrightarrow{AB}.$$

Since (compare 1. 14) a point X of π_1 is uniquely determined by the two distances $r(X, A)$ and $r(X, C)$, γ must carry X into the unique point $X' \subset \pi_1$ with $r(X', B) = r(X, A)$ and $r(X', D) = r(X, C)$. This proves our statement.

A translation γ along g, therefore, is uniquely determined by the fact that it carries the first of a pair of points A, B on g into the second. We designate γ by $(A \rightarrow B)$. It follows that

$$(A \rightarrow B)(B \rightarrow C) = (A \rightarrow C) \qquad (\text{first } (A \rightarrow B))$$

and since the transformations of g into itself induced by translations along g are commutative, the translations along g are commutative. We now assume that to each pair of points A, B on g the translation $(A \rightarrow B)$ exists. Then these translations form an Abelian group G. Hereafter, we indicate the assumption of the presence of G by saying that *all translations along* g *exist.* We first notice some simple consequences of the existence of G.

(2. 1) *Translations different from the identity have no fixed points.*

If P remained fixed under $(A_0 \rightarrow A_1) \neq 1$, P would also be fixed under the positive powers $(A_0 \rightarrow A_2)$, $(A_0 \rightarrow A_3), \cdots$ of $(A_0 \rightarrow A_1)$. We should have

$$r(A_0, P) = r(A_1, P) = r(A_2, P) = \cdots;$$

but
$$r(A_0, A_\nu) = \nu \cdot r(A_0, A_1) \rightarrow \infty.$$

(2. 2) *Each point* P *has a unique foot on* g.

Assume that $P \subset \pi_1$ has two different feet F_1 and F_2 on g. The proof of (1. 5) shows that we can choose a point Q in π_1 with a unique foot F on G and $r(Q, g) > r(P, F_1) + r(P, F_2)$. Let F' be any point between F_1 and F_2.

$(F \to F')$ transforms Q into a point Q' which has F' as unique foot on g. Moreover Q' cannot lie in the interior of the triangle PF_1F_2. Hence $\overline{QF'}$ must intersect $\overline{F_1P} + \overline{PF_2}$ in a point R, which would have two different feet on g, contrary to page 230.

(2.3) *The points of* Σ *having the same point* F *of* g *as foot form a s. l., which we call the perpendicular to* g *in* F.

This is always true as soon as each point of Σ has exactly one foot on g, i. e. the existence of G is unessential. For if $P_1 \subset \pi_1$ and $P_2 \subset \pi_2$, both have F as foot, $\overline{P_1P_2}$ must intersect g in F. If $Q \neq F$ were the intersection, we should have

$$r(P_1, F) + r(F, P_2) > r(P_1, P_2) = r(P_1, Q) + r(Q, P_2)$$

and either

$$r(P_1, F) > r(P_1, Q) \quad \text{or} \quad r(F, P_2) > r(Q, P_2).$$

If, now, P_3 is any other point, say in π_1, with F as foot, then for the same reason P_3 must be on $\underline{P_2Q}$ $(= \underline{P_1Q})$.

As an immediate consequence we have

(2.4) $(A \to B), A, B \subset g$, *transforms the perpendicular* p *to* g *at* A *into the perpendicular* q *to* g *at* B.

The point on p in π_1 which has distance r from A is carried into the point in π_1 on q which has distance r from B. We consider all the points which have distance r from g. They form two curves c_r^1, c_r^2, the one in π_1, the other in π_2. c_r^1 and c_r^2 are equidistant from g and are transformed into themselves by each translation along g. We now prove the important fact:

(2.5) *The curves* c_r^1 *and* c_r^2 *are convex; that is, a suitable one of the sides of* $c_r^1(c_r^2)$ *(called the interior of* $c_r^1(c_r^2)$*) together with* $c_r^1(c_r^2)$ *has the property of containing a s. l. segment* $\overline{AB}$ *if it contains* A *and* B.

We consider c_r^1 for some $r > 0$. If c_r^1 contains a s. l. segment it is a s. l., since it is transformed into itself by each motion of G.

We therefore assume that c_r^1 contains no s. l. segment. Let A', B' be any points of c_r^1, $\widehat{A'_1B'}$ the corresponding arc of c_r^1. If $\underline{A'B'}$ intersects $\widehat{A'B'}$ in more points than A' and B' we can find a subarc $\widehat{AB}$ of $\widehat{A'B'}$ such that $\widehat{AB} \cdot \underline{AB} = A + B$. Let π be the half plane determined by $\underline{AB}$ which contains

$\overset{\frown}{AB} - A - B$, and D any point of $\underrightarrow{AB} - \overline{AB}$. The perpendicular a to g through D does not intersect $\overset{\frown}{AB}$, because each perpendicular intersects c_r^1 exactly once and $\overset{\frown}{AB}$ is met by the perpendiculars through the points of $\overline{AB}$. Let s be a ray of π^* issuing from D. We vary s continuously from the position $a \cdot \pi^*$ towards $\underrightarrow{DB}$. $\overset{\frown}{AB}$ being bounded, there exists a first ray s^0 having common points with $\overset{\frown}{AB}$, $s^0 \neq DB$. Let T be any point of $\overset{\frown}{AB} \cdot s^0$. Using again the fact that each perpendicular to g intersects c_r^1 exactly once one sees that there exists a circular disc with center T such that the (open) half λ_T, which is on the same side of s^0 as $a \cdot \pi^*$ has no common points with c_r^1.

Let T_1 be any point of c_r^1. The translation along g carrying T into T_1 transforms λ_T into a congruent semi-circular disc with center T_1 disjoined from c_r^1 and on the same side of c_r^1 as λ_T. Calling this side the exterior and the other the interior of c_r^1, a theorem of Tietze[11] shows that the interior of c_r^1 plus c_r^1 is convex. It is true that Tietze assumes the metric to be Euclidean, but the simple proof for Tietze's theorem given by Reinhardt[12] can be carried over to our case without any change whatsoever.

From this proof it follows that at each point of c_r^1 there exists a supporting line (this can be proved quite generally for convex curves in arbitrary two-dimensional SL spaces). For, since c_r^1 contains no segment and the s. l. $\bar{s}$ bearing s^0 certainly contains an exterior point of c_r^1, the s. l. $\bar{s}$ must be, except for T, completely in the exterior of c_r^1. By translations of $\bar{s}$ along g we get supporting lines at each point of c_r^1. It is also easy to see that $\bar{s}$ is the only supporting line of c_r^1 at T such that c_r^1 has at each point a unique tangent. But we do not need this later on. We determine now which side of c_r^1 is the interior by proving

(2.6) *The domain bounded by* c_r^1 (c_r^2) *and* g *is convex.*

Using T and $\bar{s}$ with the same meaning as previously, at least one, say $\bar{s}'$, of the two rays on $\bar{s}$ issuing from T does not intersect g. Let the perpendicular to g through T cut g at Q and let Q_1 be a point on g on the same ("right") side of the perpendicular as $\bar{s}'$. If $r(Q, Q_1) > 0$ is sufficiently small the translation $(Q \rightarrow Q_1)$ transforms $\bar{s}$ into an s. l. $\bar{s}_1$ intersecting $\bar{s}$ in a point L of $\bar{s}'$. If T_1 is the image of T under $(Q \rightarrow Q_1)$, the ray $\underrightarrow{LT_1}$ contains the image of $\bar{s}'$. c_r^1 intersects all perpendiculars to g. Therefore, if the assertion were not true, i. e. if $\bar{s}$ and $\bar{s}_1$ were except for T respectively T_1 between c_r^1

---

[11] In [12].

[12] In [10].

and g, the ray $\underrightarrow{LT_1}$ would have to intersect all perpendiculars to g on the right side of the one through L. Let X_1 traverse $\underrightarrow{LT_1}$ and let X be the point where the perpendicular to g through X_1 intersects $\bar{s}'$. Since each curve c_r^1 is cut at most twice by $\bar{s}_1$ and $\bar{s}$, and these s. l. are between c_r^1 and g, the numbers $r(X_1, g)$ and $r(X, g)$ should converge to certain limits r_0^1 and r_0 with $0 \leq r_0^1 \leq r$ and $0 \leq r_0 \leq r$. Then we must have $r_0 = r'_0$ since $\bar{s}_1$ is the image of $\bar{s}$ under $(Q \rightarrow Q_1)$. Hence we should have $r(X, X') \rightarrow 0$ which contradicts our Lemma (1.9).

We conclude from (2.6) that an s. l. h intersecting g in a point A meets each curve c_r^1 or c_r^2 at most once and is therefore transformed by any translation $(A \rightarrow A')$ of G into an s. l. h' not intersecting h. For, if $h \cdot h' = S$, $(A \rightarrow A')$ would carry S into a point S' on h', A', S, S' would be on h' and S and S' would be on the same curve c_r^1 or c_r^2. We see, furthermore, that if X traverses one of the rays of L determined by A in one direction, $r(X, g)$ increases. Distinguishing $\overrightarrow{g}$ and $\overleftarrow{g}$ one finds: If Y traverses an asymptote a to $\overrightarrow{g}$ in such a direction, that the foot F_Y of Y traverses $\overrightarrow{g}$, the distance $r(Y, F_Y)$ either decreases monotonically in the strict sense or is constant for all Y on a. For, if $r(Y, F_Y)$ is constant on a certain segment of a this segment belongs to a curve c_r^i; then a coincides with this c_r^i. This leads to the following theorem:

(2.7) *A curve c_r^i is a shortest line if, and only if, it is parallel to g.*

For, let $p \neq g$ be a parallel to g in π_1. Our last statement shows that $r(Y, g)$ decreases or remains constant if Y traverses p in either direction; hence $r(Y, g)$ is constant and p is a curve c_r^1.

The converse is a little more involved. We show first: If c_r^1 is a s. l. and $(A \rightarrow B)$ transforms the point $\bar{A}$ of c_r^1 into the point $\bar{B}$, then

(2.8) $$r(A, B) = r(\bar{A}, \bar{B}).$$

Let A_0, B_0 be the feet of $\bar{A}, \bar{B}$ on g. We have

$$r(A_0, B_0) = r(A, B).$$

Call B_0^{n-1}, $\bar{B}^{n-1}$ the images of B_0 respectively B under $(A \rightarrow B)^n$. The points $\bar{B}, \bar{B}^1, \cdots, \bar{B}^{n-1}$ are on c_r^1. We have

$$r(A_0, B_0^{n-1}) = n \cdot r(A_0, B_0)$$
$$r(\bar{A}, \bar{B}^{n-1}) = n \cdot r(\bar{A}, \bar{B})$$

hence

$$n \cdot | r(A_0, B_0) - r(\bar{A}, \bar{B}) | = | r(A_0, B_0^{n-1}) - r(\bar{A}, \bar{B}^{n-1}) |$$
$$< r(\bar{A}, A_0) + r(\bar{B}^{n-1}, B_0) = 2r.$$

The same argument gives the more general fact: If $(A \rightarrow B)$ transforms $\bar{A}$ into $\bar{B}$, then

$$r(\bar{A}, \bar{B}) \geq r(A, B).$$

For, considering the images of $\bar{B}_0$ and B_0 under $(A \rightarrow B)$ we find, as above,

$$n \cdot r(\bar{A}, \bar{B}) + 2r \geq n \cdot r(A, B).$$

We remark, furthermore, that the metric characterization (1.7) of the limit spheres yields:

(2.9) *$(A \rightarrow B)$ transforms the limit circle $L(A, \overrightarrow{g})$ into $L(B, \overrightarrow{g})$ and $L(A, \overleftarrow{g})$ into $L(B, \overleftarrow{g})$.* With the help of (2.8) and (2.9) we prove (2.7) as follows. If c_r^1 were an s. l. but not an asymptote, for instance, to $\overrightarrow{g}$, the asymptote a to $\overrightarrow{g}$ through a point P of c_r^1 would be different from c_r^1. The limit circle to $\overrightarrow{g}$ through P may cut g in P_0. Let $(P_0 \rightarrow Q_0)$ be any translation in the direction $\overleftarrow{g}$. $(P_0 \rightarrow Q_0)$ transforms $L(P_0, \overrightarrow{g})$ into $L(Q_0, \overrightarrow{g})$. Designating the intersections of $L(Q_0, \overrightarrow{g})$ with c_r^1 and a by Q and Q', we would have $r(P, Q') = r(P_0, Q_0)$ on account of (1.8) and $r(P, Q) = r(P_0, Q_0)$ as a consequence of (2.8), in contradiction to (1.10).

We see that a translation $(A \rightarrow B)$ along g is also a translation $(\bar{A} \rightarrow \bar{B})$ along c_r^1, with $r(A, B) = r(\bar{A}, \bar{B})$, if c_r^1 is an s. l. and $\bar{A} < c_r^1$. (2.6) shows that then the curves c_r^1 with $0 < r < r_0$ are also s. l. Putting our results together we find:

(2.10) *If $c_{r_0}^1$, $r_0 > 0$, is an s. l., then the curves c_r^1 with $0 \leq r \leq r_0$ are also* s. l.; *all these c_r^1 are parallel to each other; and a translation $(A \rightarrow B)$ along g is at the same time a translation $(A' \rightarrow B')$ along any of the c_r^1 with $r(A, B) = r(A', B')$. An s. l. h intersecting g intersects all c_r^1 for $0 \leq r \leq r_0$. If h' is the transform of h under $(A \rightarrow B)$, $A, B \subset g$, the* s. l. *h and h' cut out segments of the same length $r(A, B)$ on all these c_r^1 and two different c_r^1 cut out segments of equal length on h and h'.*

This could lead to the conjecture that if through each point of the plane a parallel to g exists (we say in this case: the parallel axiom holds with respect to g), the parallel axiom must hold with respect to each s. l. But we show by an example that this is not necessarily so, not even in Desarguesian geometry. We introduce a metric in the part $|y| < 1$ of the (x, y)-plane by putting

$$r((x_1, y_1), (x_2, y_2)) = |x_1 - x_2| + \left| \log \frac{(y_1 - 1)(y_2 + 1)}{(y_1 + 1)(y_2 - 1)} \right| + \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2}.$$

One sees easily that $r(\;)$ defines an SL space and that the parts of the Euclidean straight lines in $|y| < 1$ are the s. l. of the metric, which therefore is Desarguesian. Furthermore, the parallel axiom holds with respect to the x-axis and all translations along the x-axis exist. We remark that the s. l. $x = \text{constant}$, $|y| < 1$, are equidistant from each other. The circles of this metric are in general not convex. A result of P. Funk[13] shows that it is impossible to introduce in $|y| < 1$ a (Desarguesian) metric with these s. l. for which the equidistant curves to each straight line are again straight lines. Section 4 of this paper implies that one cannot find both a metric with these s. l. and translations along two s. l. which are neither parallel nor asymptotes to each other.

We have seen that *the existence of G, the Desarguesian character of the metric, and the parallel axiom with respect to g do not imply the validity of the parallel axiom with respect to each s. l.* On the other hand, we are going to show by another example that the existence of G and the validity of the parallel axiom for each s. l. do not imply the Desarguesian character of the metric.

For, let Σ be the whole Euclidean plane and let g be the x-axis. We introduce a Minkowskian metric $r_1(P, Q)$ in $\pi_1 + g$ $(y \geqq 0)$ and a different one, $r_2(P, Q)$ in $\pi_2 + g (y \leqq 0)$ but so that the diameters parallel to g of the unit circles both have the Euclidean length 1. Then one has ($e(\;)$ is the Euclidean distance) $e(P, Q) = r_1(P, Q) = r_2(P, Q)$ for $P, Q \subset g$. Let $P_1 = (x_1, y_1)$, $y_1 > 0$ be any point π_1 and $P_2 = (x_2, y_2)$, $y_2 < 0$ any point of π_2. Then the functions $r_1(P_1, (x, 0))$ and $r_2(P_2, (x, 0))$ are convex; hence their sum is convex and there exists a uniquely determined point $X_0 \subset g$, such that

$$r(P_1, P_2) = \min_X [r_1(P_1, (x, 0)) + r_2(P_2, (x, 0))] = r(P_1, X_0) + r(P_2, X_0).$$

Putting

$$\begin{aligned} r(P, Q) &= r_1(P, Q) \quad \text{if} \quad P, Q \subset \pi_1 + g \\ r(P, Q) &= r_2(P, Q) \quad \text{if} \quad P, Q \subset \pi_2 + g \end{aligned} \tag{2.11}$$

---

[13] In [4].

defines $r(P, Q)$ for all points of the plane. One sees easily that $r(P, Q)$ satisfies all our conditions. The s. l. are the straight lines $y = \text{const.}$ and straight lines which are broken at a point of the x-axis. Parallel rays in $\pi_1 + g$ have parallel continuations in $\pi_2 + g$; therefore *the parallel axiom holds for each* s. l. *All translations along each line $y = const.$ exist. But the theorem of Desargues is not true,* if the circles of the two Minkowskian geometries have no special relations to each other, for instance, if the circles in π_1 are ellipses and those in π_2 are not.

Returning to the general case, suppose that there exists a last $r \geqq 0$, say r_0, such that $c_{r_0}{}^1$ is an s. l. g_1, and let π^* be that half plane determined by g_1, which is contained in π_1. Then through each point of π^* there pass two different asymptotes to g_1. Let a be an asymptote to $\overleftarrow{g}_1$. From (2. 9) it follows that the images of a under the translations of G are again asymptotes to $\overleftarrow{g}_1$, and (2. 6) shows that a translation γ in the direction $\overrightarrow{g}$ carries c into an s. l. between a and g_1. Hence the positive powers of γ must carry a into a sequence of asymptotes to $\overleftarrow{g}_1$, which, according to (1. 11), converges to an asymptote $\bar{a}$ to $\overleftarrow{g}_1$. $\bar{a}$ must remain invariant under all translations of G; therefore $\bar{a}$ is a curve $c_r{}^1$; and since $c_{r_0}{}^1 = g_1$ is the last curve $c_r{}^1$ which is an s. l., we must have $\bar{a} = g_1$. This implies that each asymptote a to g_1 in π^* has distance 0 from g_1, and that $r(x, g_1) \rightarrow \infty$ if x traverses a in the direction $\overrightarrow{a}$. From (1. 15) and (1. 16) we have that $\overleftarrow{a}$ is also an asymptote to $\overleftarrow{g}_1$ and to each other asymptote to $\overleftarrow{g}_1$ in π^*. Let h be any s. l. in π^* which is no asymptote to g. Orienting h in the "same" way as g, one sees easily that there exists an s. l. a_1 between h and g which is an asymptote to $\overrightarrow{h}$ and $\overleftarrow{g}_1$ and an s. l. a_2, which is an asymptote to $\overleftarrow{h}$ and $\overrightarrow{g}_1$. Hence, if y tends to infinity on h in either direction, one has $r(y, g_1) \rightarrow \infty$. There exists exactly one pair of points $S \subset h$ and $R \subset g_1$ such that $r(R, S) = r(h, g_1)$. S is determined as the point where a curve $c_r{}^1$ touches h and R in the foot of S on g_1 and inversely.

The first question which arises in this connection is if it is actually possible that $g_1 \neq g$ *or* $0 < r_0 < \infty$. *In a Desarguesian geometry it certainly can not happen.* For, if one of two Euclidean straight line segments (unbounded ones admitted) is parallel to another in our sense, both of them must be whole straight lines. The only convex domains in the Euclidean plane which contain a whole straight line are the whole plane, a half-plane, and the part of the plane between two parallel lines. *But there are non-Desarguesian geometries with* $0 < r_0 < \infty$.

In order to construct such a metric, let ζ be the interior of the unit circle of the (x, y)-plane. Introduce in ζ a hyperbolic metric $h(P, Q)$ in such a way that the straight-line segments in ζ become the s. l. Let ζ_1 be the part $y \leq 0$, ζ_2 the part $y \geq 0$ of ζ, and π^* the half-plane $y \leq 0$. We map π^* topologically on ζ_2 by associating $O = (0, 0)$ with itself and a point $Q \neq 0$ to that point $\bar{Q}$ on $\underrightarrow{OQ}$ for which

$$e(O, Q) = h(O, \bar{Q}) \quad (e(\ ) \text{ is the Euclidean distance}).$$

If $(\bar{P}, \bar{Q})$ are any points in ζ_2 and P, Q the points in π^* to which they correspond, we put

$$r(\bar{P}, \bar{Q}) = e(P, Q).$$

In this way we have introduced in ζ_2 a Euclidean metric (for which the s. l. are in general no Euclidean straight line segments) which on $0 < x < 1$, $y = 0$ coincides with the hyperbolic metric. For two points in ζ_1 we put

$$r(\bar{P}, \bar{Q}) = h(\bar{P}, \bar{Q}).$$

By applying the same method which led to the combination of the two different Minkowskian geometries, we define $r(P, Q)$ also for $P < \zeta_1$, $Q < \zeta_2$. We thus get a metric in ζ, which makes ζ an SL space, with translations along the s. l. $g_1 : x = 0$. Taking any parallel $g \neq g_1$ to g_1 in ζ_2 one sees easily that these s. l. g and g_1 satisfy the assumptions of our previous considerations.

3. Geometries with translations along an s. l. g and its asymptotes. The last example also shows that the existence of all translations along g_1 (or g) does not imply the existence of translations along an asymptote to g_1 (or g) which is not a parallel.

Our considerations of the general case (compare p. 242) did indicate that by translations along g_1 (or g) any asymptote $\overrightarrow{a} \neq \overrightarrow{g_1}$ to $\overrightarrow{g_1}$ in π^* can be carried into any other one. If there exist also translations along a, we conclude from this that there exist two different asymptotes to $\overrightarrow{g_1}$ through each point $P \not\subset g_1$. For, if $g \neq g_1$, g would be an asymptote to $\overrightarrow{a}$, since g is an asymptote to $\overrightarrow{g_1}$ and $\overrightarrow{g_1}$ to $\overrightarrow{a}$ (compare (1.15)). Then $r(a, g) = 0$, but $r(a, g) \geq r(g_1, g) > 0$. From $r(a, b) = 0$ for each asymptote $\overrightarrow{b}$ to $\overrightarrow{a}$ and

$r(g, c) = 0$ for each asymptote $\overrightarrow{c}$ to $\overrightarrow{g} = \overrightarrow{g}_1$, one sees that the asymptotes to $\overrightarrow{g}$ are asymptotes to each other; and, since by a translation along $a^{(g)}$ any asymptote to g in π_2 (π_1) can be transformed into $g^{(a)}$ there exist translations along each asymptote to g. We have found

(3.1) *If $\overrightarrow{a}$ is an asymptote to $\overrightarrow{g}$ not parallel to g and if all translations exist along g as well as along a, then all asymptotes to $\overrightarrow{g}$ are asymptotes to each other and all translations along each of these s. l. exist. If b is any of them, then the two asymptotes to b through any point P not on b are different.*

We shall now study in a little more detail the properties of a metric satisfying the assumption of (3.1). Let π_1 and π_2 again be the two half planes determined by g. At first all considerations will refer only to one of these half planes, say π_1. We therefore put $P + L(P, \overleftarrow{g})\pi_1 = L_1(P, \overleftarrow{g})$. We again designate by c_r^1 the curve equidistant from g in π_1 with $r(c_r^1, g) = r$. (2.9) shows that the arcs of the different limit circles $L(P, \overleftarrow{g})$ between any two fixed curves $c_{r_1}^1$ and $c_{r_2}^1$ are congruent. The availability of the translations along the asymptotes to g allows a considerable strengthening of this statement. Each arc $\overset{\frown}{AB}$ on $L_1(P, \overleftarrow{g})$ is congruent to an arc on any $L_1(Q, \overleftarrow{g})$ starting at g and therefore also to an arc starting at an arbitrary point. (We say the arc $\overset{\frown}{AB}$ starts at A if A is on $L_1(P, \overleftarrow{g})$ between B and P.) To prove this we draw the asymptote a to $\overleftarrow{g}$ through B and the curve through A equidistant from a. Since $r(a, g) = 0$, this curve intersects g at a point A'. We draw $L(A', \overleftarrow{g})$ and put $a \cdot L(A', \overleftarrow{g}) = B'$. The curve c_r^1 through B' may intersect $L_1(Q, \overleftarrow{g})$ at Q'. We say that $\overset{\frown}{QQ'}$ is congruent to $\overset{\frown}{AB}$ (and write $\overset{\frown}{QQ'} \cong \overset{\frown}{AB}$). Since $(Q \rightarrow A')$ transforms $\overset{\frown}{QQ'}$ into $\overset{\frown}{A'B'}$, we have $\overset{\frown}{QQ'} \cong \overset{\frown}{A'B'}$. The translation $(B \rightarrow B')$ along a transforms $L(P, \overleftarrow{g})$ into the limit circle through B' to the image $\overleftarrow{g}'$ of $\overleftarrow{g}$ under $(B \rightarrow B')$. $\overleftarrow{g}'$ is an asymptote to $\overleftarrow{g}$ in π_2. The limit circles to $\overleftarrow{g}'$ and $\overleftarrow{g}$ coincide in π_1 on account of (1.13), hence $\overset{\frown}{QQ'} = \overset{\frown}{A'B'} = \overset{\frown}{AB}$. It follows from this that on each arc $\overset{\frown}{AB}$ of $L_1(P, \overleftarrow{g})$, there exists a point C (the center of AB) with $\overset{\frown}{AC} \cong \overset{\frown}{CB}$ and more generally points $C_1, \cdots, C_{n-1}$ with

$$\widehat{AC_1} \cong \widehat{C_1C_2} \cong \cdots \cong \widehat{C_{n-1}B}.$$

We then write

$$\widehat{PQ} = \widehat{A'B'} = 2\widehat{AC} = 2\widehat{CB}$$
$$\widehat{AB} = n\,\widehat{C_\nu C_{\nu+1}} = \frac{n}{m}\,\widehat{C_\nu C_{\nu+m}}, \text{ and so forth,}$$

and extend this notation to arbitrary non-negative real numbers by a limit process. A consideration similar to the proof of (2.5) shows that the limit circle $L_1(P, \overleftarrow{g})$ are convex and therefore rectifiable, and that congruent pieces have the same length. If one does not wish to use this fact, he may associate the number 1 to an arbitrary arc $\widehat{AB}$, $A \neq B$, of any $L_1(P, g)$ and real

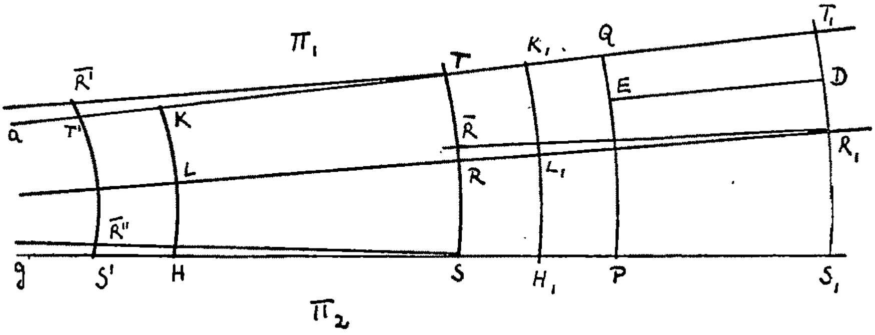


Fig. 1.

numbers to other arcs according to our above prescription. We use $\widehat{AB}$ both for the arc and its length.

We intend to determine the length of an arc $\widehat{YX}$, where X traverses g and Y a fixed asymptote a to $\overleftarrow{g}$ as a function of the distances on g (see Fig. 1). Let $\widehat{ST}$, $S \subset g$, $T \subset a$ be any such arc; draw the curve through T equidistant from g and the one through S equidistant from a. They may intersect in R_1. Let the limit circle through R_1 to g intersect g in S_1 and a in T_1. Then, as before, $\widehat{S_1R_1} \cong \widehat{R_1T_1}$. The asymptote to $\overleftarrow{g}$ through R_1 intersects $\widehat{ST}$ in its center $\bar{R}$. If the intersection were different from $\bar{R}$, say $R \subset S\bar{R} - \bar{R}$, then $(S_1 \rightarrow S)$ would transform $\widehat{S_1R_1}$ into $\widehat{ST}$, $\underline{R_1R}$ into a, and hence $\underrightarrow{R_1\bar{R}}$ into a ray $\underrightarrow{T\bar{R}'}$, where $\bar{R}'$ would be on the limit circle $L(S', \overrightarrow{g})$ through the image S' of S under $(S_1 \rightarrow S)$. Put $L(S', \overrightarrow{g}) = T'$. Then $(T_1 \rightarrow T)$ transforms $\widehat{T_1R_1}$

into $\overset{\frown}{TS}$, $\underline{R_1R}$ into g, $\overset{\frown}{TR}$ into $\overset{\frown}{T'S'}$, and $\underrightarrow{R_1\bar{R}}$ into a ray $\underrightarrow{S\bar{R}''}$, where $\bar{R}''$ is an interior point of $\overset{\frown}{T'S'}$. But then we should have

$$\overset{\frown}{T'\bar{R}''} \cong \overset{\frown}{T\bar{R}} \cong \overset{\frown}{\bar{R}S} \cong \overset{\frown}{\bar{R}'S'}$$

whereas $\overset{\frown}{T'\bar{R}''}$ is a proper subarc of $\overset{\frown}{\bar{R}'S'}$.

Since $\overset{\frown}{S_1R_1} = 2\overset{\frown}{RS}$, we see the s. l. connecting the centers of $\overset{\frown}{S_1R_1}$ and $\overset{\frown}{SR}$ is an asymptote to g and, since the arcs $\overset{\frown}{R_1T_1}$ and $\overset{\frown}{RT}$ also belong to limit circles to $\underrightarrow{R_1R}$, we see that the s. l. connecting the centers of $\overset{\frown}{R_1T_1}$ and $\overset{\frown}{RT}$ is also an asymptote to $\overleftarrow{g}$. We thus find:

Let the points $B_1 \subset \overset{\frown}{S_1T_1}$ and $A \subset \overset{\frown}{ST}$ be chosen in such manner that

$$\overset{\frown}{SA} = \lambda\overset{\frown}{ST}, \qquad \overset{\frown}{S_1A_1} = \lambda\overset{\frown}{S_1T_1}\ (= 2\lambda\overset{\frown}{ST}), \qquad 0 < \lambda < 1.$$

Then $\underline{AB}$ is an asymptote to g.

Furthermore our considerations show that

$$\overset{\frown}{S'T'} = \tfrac{1}{2}\overset{\frown}{ST} = \tfrac{1}{4}\overset{\frown}{S_1T_1}, \qquad r(S', S) = r(S, S_1).$$

More generally, if S_n and $S^{(n)}$ are chosen in such manner that

$$S_n \subset \underrightarrow{S'S} \text{ and } r(S, S_n) = nr(S', S) = nr(S, S_1)$$
$$S^{(n)} \subset \underrightarrow{SS'} \text{ and } r(S, S^{(n)}) = nr(S', S)$$

and if one puts $L(S^{(n)}, \overleftarrow{g}) \cdot a = T^{(n)}$, $L(S_n, \overleftarrow{g}) \cdot a = T_n$ one has

$$\overset{\frown}{S^{(n)}T^{(n)}} = \frac{1}{2^n}\overset{\frown}{ST}, \qquad \overset{\frown}{S_nT_n} = 2^n\overset{\frown}{ST}.$$

Now choose any point H between S' and S and put $a \cdot L(H, \overleftarrow{g}) = K$. Let H_1 be the point on g for which, with $L(H_1, \overleftarrow{g}) \cdot a = K_1$, one has $\overset{\frown}{H_1K_1} = 2\overset{\frown}{HK}$. We assert that

$$r(H, H_1) = r(S, S_1).$$

We consider the points H_n on $\underrightarrow{HH_1}$ with $r(H_n, H) = nr(H_1, H)$ and put $a \cdot L(H_n, \overleftarrow{g}) = K_n$. If, for instance, one had $r(H, H_1) > r(S, S_1)$ we should have

$$r(S, S_\nu) < r(S, H_\nu)$$

for large ν; hence with an obvious signification of the sign $>$

$$2^\nu \overset{\frown}{HK} = \overset{\frown}{H_\nu K_\nu} > \overset{\frown}{S_\nu T_\nu} = 2^\nu \overset{\frown}{ST}$$

in contradiction to the fact that $\overset{\frown}{ST} > \overset{\frown}{HK}$.

Up to this point we have found:

(3.2) *There exists a number* $\sigma > 0$ *such that, for an arbitrary asymptote* b *to* $\overleftarrow{g}$ *in* π_1 *and for an arbitrary pair of points* $A < B$ *on* $\overleftarrow{g}$ *with* $r(A, B) = \sigma$, *the arc of* $L(A, \overleftarrow{g})$ *between* b *and* g *is congruent to the half of the arc of* $L(B, \overleftarrow{g})$ *between* g *and* b.

We now prove:

(3.3) *The points* Z *of all arcs* $\overset{\frown}{XY}$ *of* $L(X, \overleftarrow{g})$ *with* $X \subset g$, $Y \subset a$ *and* $\overset{\frown}{XZ} = \lambda XY$, $0 < \lambda < 1$, *form an asymptote to* g.

As in a previous proof, it is sufficient to show this for $\lambda = \frac{1}{2}$. Using the same notations as before and calling L, L_1 the centers of $\overset{\frown}{HK}$, respectively $\overset{\frown}{H_1K_1}$, we conclude from $r(H, H_1) = r(S, S_1) = \sigma$ that $(H \rightarrow H_1) = (S \rightarrow S_1)$. But $(H \rightarrow H_1)$ transforms a into the asymptote to $\overleftarrow{g}$ through that point M of $\overset{\frown}{H_1K_1}$ for which $\overset{\frown}{H_1M} \cong \overset{\frown}{HK}$; hence $M = L_1$. $(S \rightarrow S_1)$ transforms a into $\underline{R_1R}$; hence $\underline{R_1R} \equiv \underline{L_1L}$. This proves the assertion.

Herewith the length of $\overset{\frown}{XY}$ can be determined as follows: Let P be the center of $\overline{SS_1}$ and put $L(P, \overleftarrow{g}) = Q$. $(S \rightarrow P)$ transforms $\overset{\frown}{ST}$ into a subarc $\overset{\frown}{PE}$ of $\overset{\frown}{PQ}$. Let the asymptote to $\overleftarrow{g}$ through E intersect $\overset{\frown}{S_1T_1}$ in D. Then $\overset{\frown}{PQ} \cong \overset{\frown}{S_1D}$ since $(S \rightarrow P)$ transforms a into $\underline{ED}$, and, on account of (3.3),

$$\frac{\overset{\frown}{ST}}{\overset{\frown}{PQ}} = \frac{\overset{\frown}{PE}}{\overset{\frown}{PQ}} = \frac{\overset{\frown}{S_1D}}{\overset{\frown}{S_1T_1}} = \frac{\overset{\frown}{PQ}}{2\overset{\frown}{ST}}$$

or

$$PQ = \sqrt{2}\, \overset{\frown}{ST}.$$

Putting

$$r(S,X) = x \text{ for } X \subset \underrightarrow{SS'}$$
$$r(S,X) = -x \text{ for } X \subset \underrightarrow{SS_1}$$

we find in this way

$$\widehat{XY} = \widehat{ST}\, e^{-x[(\log 2)/\sigma]} \tag{3.4}$$

which (except for the normalization $\sigma = \log 2$) is the same formula as in hyperbolic geometry.[14]

We have derived (3.4) using only a part of the congruence axioms. We show, indeed, by an example that *a geometry with all translations along two non-parallel asymptotes is not necessarily hyperbolic, not even Desarguesian.*

We again introduce in the domain $x^2 + y^2 \leq 1$ of the (x, y)-plane the hyperbolic metric $h(P, Q)$ of —. Designate by h the broken line consisting of the two hyperbolic rays issuing from $(0, 0)$ and ending at $(0, -1)$ respectively $(\sqrt{2}, \sqrt{2})$. We define distances for the points P', Q' on h as follows: Call g the hyperbolic s. l. $y = 0$, $|x| < 1$ and let c_r^1, c_r^2 be the hyperbolic curves equidistant from g in $y < 0$ respectively $y > 0$. Each of these curves c_r^1, c_r^2 intersects h exactly once. We draw the curves passing through P' and Q'. If both of them are curves c_r^1 (c_r^2), say c_r^1, $c_{r_2}^1$ $(c_{r_1}^2, c_{r_2}^2)$, we put $r(P', Q') = |r_1 - r_2|$. If one is the curve $c_{r_1}^1$, the other $c_{r_2}^2$ we put $r(P', Q') = r_1 + r_2$. Call $\bar{G}$ the smallest group of hyperbolic motions containing the translations along the s. l. ending at $(-1, 0)$. $\bar{G}$ can be generated by the translations along g and the hyperbolic rotations around the point $(-1, 0)$. By a suitable transformation of $\bar{G}$ any pair of points P, Q, such that $\underline{PQ}$ (by $\underline{XY}$, respectively $\overline{XY}$, we designate during the discussion of this example the Euclidean straight line, respectively segment, connecting X and Y) does not pass through $(-1, 0)$, can be transformed into exactly one pair of points P', Q' on h. We put

$$r(P, Q) = r(P', Q'). \tag{3.5a}$$

For pairs of points P, Q, where $\underline{PQ}$ passes through $(-1, 0)$ we put

$$r(P, Q) = h(P, Q). \tag{3.5b}$$

It is easy, but a little tedious, to confirm that the function $r(P, Q)$ defines an SL space. The s. l. of this metric arc h and its transforms under $\bar{G}$ and the hyperbolic s. l. issuing from $(-1, 0)$. Evidently $r(P, Q)$ remains invariant under all transformations of G. The theorem of Desargues does not hold.

---

[14] See for instance [6], p. 55.

To find a configuration confirming this statement (comp. Fig. 2), we take the Euclidean ellipse e with center at $(-\frac{1}{4}, 0)$ having third order contact with the unit circle at $(-1, 0)$. e is a limit circle in both our and the hyperbolic metric:[15] Take a point A on e in $y > 0$ such that the hyperbolic tangent to e at A intersects g at a point B_{10}. A divides this tangent into two rays of which the one containing B_{10} is a part of an s. l., h_1, of our metric. Choose g_2 issuing from $(-1, 0)$ such that it intersects h_1 in a point B_{12}. Take now two hyperbolic rays issuing from A and enclosing a sufficiently small angle α which contains h_1 in its interior. Let these rays intersect g_i in B_{0i} and B_{2i}, $(i = 0, 2;$

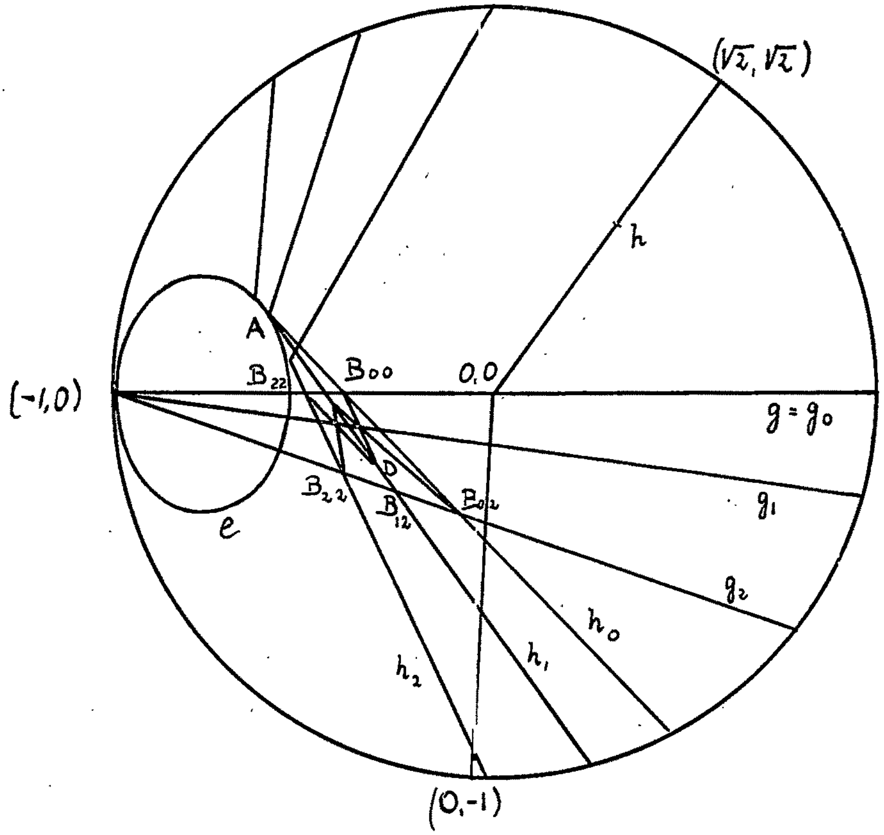


Fig. 2.

---

[15] See [13], p. 356.

$g_0 = g$) where B_{2i} is to be understood to be between $(-1, 0)$ and B_{0i}. Then the hyperbolic rays $\underrightarrow{AB_{00}}$ and $\underrightarrow{B_{20}B_{22}}$ are also rays of our metric. Let h_0 and h_2 be the s. l. of our metric containing $\underrightarrow{AB_{00}}$ and $\underrightarrow{B_{20}B_{22}}$, respectively. Then h_0 and h_1 pass through A but h_2 does not. We choose D and E on $\overline{B_{10}B_{12}}$ ($h_1 g_i = B_{1i}$) such that D is between E and B_{12}. If α is sufficiently small, the hyperbolic segments $\overline{B_{00}D}$, $\overline{B_{20}D}$, $\overline{B_{02}E}$ and $\overline{B_{22}E}$ are also segments of our metric. The points $\overline{B_{00}D} \cdot \overline{B_{02}E}$ and $\overline{B_{20}D} \cdot \overline{B_{22}E}$ together with $(-1, 0)$ lie on a hyperbolic s. l. g_1, which is also an s. l. of our metric. Hence corresponding sides of the triangles $B_{00}DB_{20}$ and $B_{02}EB_{22}$ intersect on g_1, but the s. l. h_0, h_1, and h_2 through corresponding vertices of these triangles do not pass through one point.

4. Geometries with translations along two s. l. which are not asymptotes to each other. We now assume that all translations along two s. l. g and h exist where h is not an asymptote to g (and hence g not to h, compare (3.1)). If h and g do not intersect, the curve which touches h and is equidistant from g intersects the curve which touches g and is equidistant from h in exactly two points. (This follows from (2.6), (2.7) and page 242). If A is one of these points, a suitable translation along g puts h into a position so that it passes through A and a translation along h does the same for g. These two s. l. through A are different and all translations along them exist. We see that we can restrict ourselves to the case where h and g have a common point 0. We are going to prove the following theorem.

(4.1) *If all translations along g and h exist, if h is not an asymptote to g, and if there is either a parallel $g' \neq g$ to g or a parallel $h' \neq h$ to h, then the metric is Minkowskian.*

Let a parallel $g' \neq g$ exist. We assume that h intersects g in a point 0; then it must also intersect g' in a point $0'$. $(0 \rightarrow 0')$ transforms g into an s. l. not intersecting g (comp. p. 239) through $0'$, hence into g'; and transforms g' into an s. l. g'' intersecting h in the image $0''$ of $0'$ under $(0 \rightarrow 0')$. Since g' is equidistant from g, g'' must be equidistant from g'; hence parallel to g'. By continuing this and by considering also the images of g' under $(0' \rightarrow 0)$, $(0' \rightarrow 0)^2, \cdots$ we find, with the help of (2.7) and (2.10), that the parallel axiom holds with respect to g, and that these parallels are equidistant from each other. Furthermore, the images of h under any two translations along g cut on h equal segments of all these parallels to g, and any two parallels to g cut out equal segments of all the images of h. If we knew that the parallel axiom holds with respect to each s. l. in Σ we could derive our theorems from these statements in few lines. The length of the following proof is due to the fact that we prove the parallel axiom and the theorem of Desargues simultaneously.

We introduce coördinates in our plane Σ. We distinguish the two rays g^+ and g^- of g issuing from 0 and the two rays h^+ and h^- of h issuing from 0. Through each point P of Σ there passes exactly one image h_x of h under a translation along h and exactly one parallel g_y to g. Put $h \cdot g_y = Y$ and $g \cdot h_x = X$. We have, as stated before,

$$r(X,0) = r(P,Y), \qquad r(Y,0) = r(P,X).$$

As coördinates X, Y of P we take

$$\begin{aligned} x &= r(X,0) \text{ if } X \subset g^+, & x &= -r(X,0) \text{ if } X \subset g^-, \\ y &= r(Y,0) \text{ if } Y \subset h^+, & y &= -r(Y,0) \text{ if } Y \subset h^-. \end{aligned}$$

We map Σ onto a Cartesian (x,y)-plane $\bar{\Sigma}$ by making points with the same coördinates correspond. To a motion of Σ composed of translations along g and h corresponds a translation of $\bar{\Sigma}$, and, since each translation in $\bar{\Sigma}$ can be composed of translations along the x- and the y-axes, there corresponds a a motion of Σ to each translation of $\bar{\Sigma}$.

We wish to show that each s. l. in Σ is mapped onto a straight line in $\bar{\Sigma}$. We first remark that, if any s. l. k in Σ is mapped onto a straight line $\bar{k}$ in $\bar{\Sigma}$, then all translations along k exist because all translations of $\bar{\Sigma}$ along $\bar{k}$ exist. Let k' be any transform of k by a translation along g. Since k and k' cut out equal segments of g and its parallels and a translation along k' transforms a parallel to g into a parallel to g, we see that under the translation along k the s. l. k' is transformed into itself. It is therefore equidistant from k and parallel to k. We have

(4. 2) *If any* s. l. k *in* Σ *is mapped onto a straight line* $\bar{k}$ *in* $\bar{\Sigma}$, *then the parallel axiom holds with respect to* k. *The images of the parallels to* k *in* Σ *are the Euclidean parallels to* $\bar{k}$ *in* $\bar{\Sigma}$. *All translations along* k *exist.*

Let e be any s. l. in Σ; $\bar{e}$ its image in $\bar{\Sigma}$. To each curve in $\bar{\Sigma}$ which can be deduced from $\bar{e}$ by a translation there corresponds an s. l. in Σ, since e can be carried into this s. l. by a motion of Σ. Therefore we know that by any translation of $\bar{\Sigma}$ the curve $\bar{e}$ is transformed into a curve which either coincides with $\bar{e}$, or intersects $\bar{e}$ in one point, or is disjoined from $\bar{e}$. If $\bar{e}$ has no intersections with one of its images without being identical with it, $\bar{e}$ is a straight line. If $\bar{e}$ has intersections with some of its images, the fact that it has at most one intersection with each of them, implies that $\bar{e}$ is a convex curve without parallel supporting lines.[16] Therefore $\bar{e}$ has two different, well defined asymptotes (in the Euclidean sense).

Now let us assume that there exists an s. l. e_0 through 0 in Σ whose image $\bar{e}_0$ is a convex curve but not a straight line. Let e_ϵ be an s. l. through 0 tending to e_0. Then the image $\bar{e}_\epsilon$ of e_ϵ tends to $\bar{e}_0$. The angle over which the directions of the supporting lines to $\bar{e}_\epsilon$ range tends to an angle containing the corresponding angle for $\bar{e}_0$.[17] Since we assume that the latter is positive, for small ϵ the former one is positive too and has a common interior direction with the latter. Then one can find chords $\bar{\sigma}_0$ and $\bar{\sigma}_\epsilon$ of $\bar{e}_0$ respectively $\bar{e}_\epsilon$ which are parallel to the common direction and have the same length (one or both of these chords may be subarcs of $\bar{e}_0$ respectively $\bar{e}_\epsilon$). By a translation of $\bar{\Sigma}$ the chord $\bar{\sigma}_\epsilon$ can be carried into $\bar{\sigma}_0$. Let $\bar{e}'$ be the image of $\bar{e}_\epsilon$ under this translation. $\bar{e}'$ and $\bar{e}_0$ have two common points; hence (since $\bar{e}'$ is the image of an s. l. e' in Σ) $\bar{e}' \equiv \bar{e}_0$. We see, then, that

(4.3) *For small ϵ, the curve $\bar{e}_\epsilon$ can be transformed into $\bar{e}_0$ by a translation of $\bar{\Sigma}$.*

We now consider all s. l. through 0 in Σ. The images of some of them are straight lines. These straight lines form a closed set. We consider one, $\bar{W}$, of the open angles which is bounded by two such straight lines $\bar{e}_1$, $\bar{e}_2$, but contains none of them. From (4.3) it follows that all the s. l. e of Σ in W are mapped onto convex curves $\bar{e}$ in $\bar{W}$, each of which can be transformed into any other by a translation of Σ. If $e \rightarrow e_1$ $(e \rightarrow e_2)$, we must have $\bar{e} \rightarrow \bar{e}_1$ $(\bar{e} \rightarrow \bar{e}_2)$. Therefore, one of the Euclidean asymptotes, $\bar{e}'_1$, of $\bar{e}$ must be parallel to $\bar{e}_1$ and the other must be parallel to $\bar{e}_2$. Since the Euclidean distance of $\bar{e}'_1$ and $\bar{e}$ vanishes, one sees from the fact that translations of $\bar{\Sigma}$ correspond to motions of Σ, that $r(e'_1, e) = 0$, also. According to (4.2) e'_1 is an s. l. and on account of (1.16), e is an asymptote to e'_1 such that e_1 and e would be two different asymptotes to e'_1 through 0. We have proved that all s. l. through 0 are mapped onto straight lines through $\bar{0}$. With the help of (4.2) and (2.10) one concludes easily that the metric is Minkowskian.

The other case, where there exists neither a parallel to g nor to h is much easier to deal with. We prove

---

[16] This can easily be shown. One finds a proof for this and other questions connected with it in [10].

[17] A proof for the analogous fact in a space of any dimension can be found in [1], p. 35.

(4.4) *If all translations along g and h exist, where h is not an asymptote to g, and if through each point of Σ not on g there pass two different asymptotes to g and through each point not on h two different asymptotes to h, then the metric is hyperbolic.*

The idea of the proof is obvious: One considers two non-intersecting s. l. g_{x_1} and g_{x_2} along which all translations exist and which are not asymptotes to each other. Then one takes a point S between these s. l. and the curves equidistant from g_{x_1} and g_{x_2} through S. They will (with a proper choice of S) intersect in a further point S'. The translation along g_{x_1} carrying S into S' followed by that along g_{x_2} carrying S' into S is a rotation around S. By keeping S fixed and varying g_{x_1} continuously, one varies the rotations around S continuously and gets in this way the full group of rotations around S. To make this a rigorous proof one has to show that g_{x_1} belongs to a continuous family of s. l. which are not asymptotes to g_{x_2} and along which all translations exist, and, furthermore, that the rotations around S really vary when g_{x_1} moves within that family.

To carry this out in detail, we prove first, that, if the s. l. g and h (of 4.4) intersect (which according to p. 250 always can be supposed to be the case), the transforms of g under the translations along h are not asymptotes to each other.

Put $g \cdot h = 0$. If by $(0 \to 0')$, $0' \subset h, g$ were transformed into the asymptote g' to $\overleftarrow{g}$ through $0'$, it would easily follow from p. 242 that each translation along h would transform g into an asymptote to $\overleftarrow{g}$. Therefore we would have the situation of section 3. $L(0, \overleftarrow{g})$ does not coincide with h; otherwise, a translation $(0 \to P')$ along g would transform $L(0, \overleftarrow{g})$ into $L(P', \overleftarrow{g})$ and, on the other hand, into an s. l. h'. Since $L(0, \overleftarrow{g})$ and $L(P', \overleftarrow{g})$ are equidistant, h and h' would be equidistant, and therefore parallel. Because $L(0, \overleftarrow{g})$ is different from h, a suitable asymptote g_1 to $\overleftarrow{g}$ would intersect $L(0, g)$ in a point Q_{11} such that the segment $\overline{Q_{11}Q_{10}}$ of g_1 between $L(0, \overleftarrow{g})$ and h has a length of the form $\sigma/2^n$, where σ has the same significance as in (3.2). In order to fix the ideas, let this segment be in the exterior of $L(0, \overleftarrow{g})$. Let the curve c_1 equidistant from h and through Q_{11} intersect g in $Q_{0,1}$. Let us introduce the point $Q_{0,2}$ on g, for which $Q_{0,1}$ is the center of $\overline{Q_{0,2}0}$, the curve c_2 equidistant from h and through $Q_{0,2}$, and the points $Q_{i,0}$ on $\underrightarrow{0Q_{1,0}}$ such that

$$r(Q_{i,0}, 0) = i \cdot r(Q_{1,0}, 0).$$

The asymptote to g through $Q_{i,0}$ may intersect c_1 in $Q_{i,1}$ and c_2 in $Q_{i,2}$. Then it follows from our assumption (that the asymptotes to g are the images of g under the translations along h) that all segments $\overline{Q_{i,j}\,Q_{i,j+1}}$ have length $\sigma/2^n$; that $L(0, \overleftarrow{g})$ passes through Q_{11} and Q_{22}; and that the limit circle to $\overleftarrow{g}$ through $Q_{i,0}$ passes through the points $Q_{i+1,1}$ and $Q_{i+2,2}$. From the signification of σ it follows, then, that $(Q_0,2 \rightarrow 0)$ carries $Q_{2,2}$ into $Q_{1,0}$, $Q_{3,2}$ into $Q_{2,0}$ and $Q_{4,2}$ into $Q_{3,0}$. By considering $(0 \rightarrow Q_{0,2})$, one sees that $Q_{1,2}$, $Q_{2,2}$, $Q_{3,2}$ must lie on one s. l.; hence it would follow from (2.5) to (2.7) that the curve c_1 must be an s. l. parallel to h, in contradiction to the assumptions of (4.4).

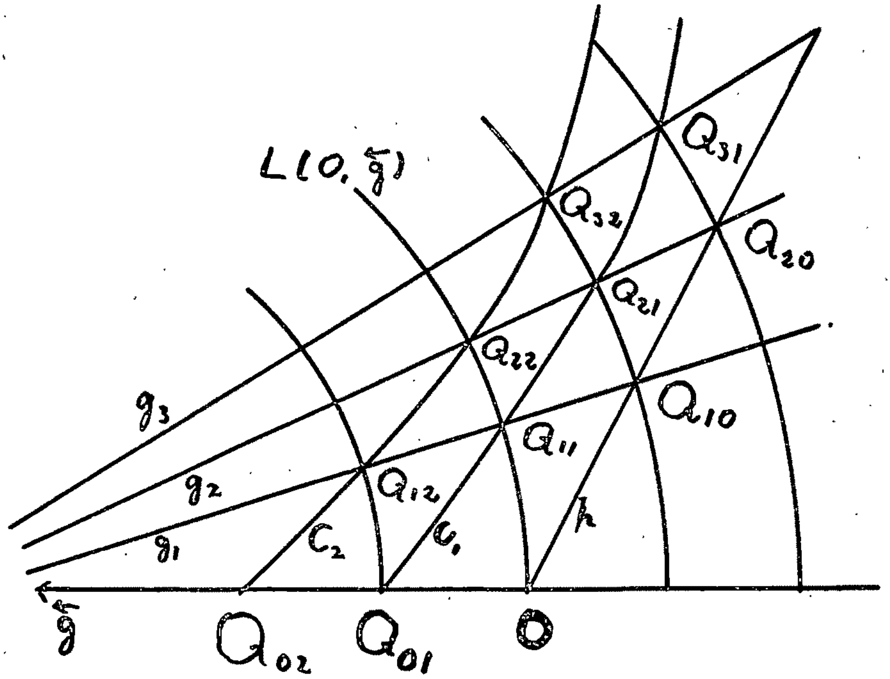


Fig. 3.

We designate the transforms of g under $(0 \rightarrow X)$, $X \subset h$, by g_x (see Fig. 4). Let X_0, X_1, X_2 be three different points on h such that X_0 is between X_1 and X_2. We choose a point S on g_{x_0} so that the curves c_1 equidistant from g_{x_1} and c_2 equidistant from g_{x_2}, through S, both cross g_{x_0} and contain in their exteriors, except for S, the same one of the two rays of g_{x_0} determined by S. This is always possible if X_0 is sufficiently near to X_1. Let S_{x_1} be the other intersection of c_1 and c_2. We consider the translations T_1 and T_2 along g_{x_1} respectively g_{x_2} which carry S into S_{x_1}. If S_{x_1} is between g_{x_0} and g_{x_2} we form $T_2T_1^{-1}$ (first T_2 then T_1^{-1}); if S_{x_1} is between g_{x_0} and g_{x_1} we form $T_1T_2^{-1}$ (for $S_{x_1} \subset g_{x_0}$ either of these motions will do). Assume the first case and put $ST_2^{-1} = K$. By $T_2T_1^{-1}$ the point K is carried into a point K_{x_1} on the side of g_{x_0} other than the one containing K. $T_2T_1^{-1}$ is a rotation with center S different from the identity. We now make the analogous transformation with g_x instead of g_{x_1}, $X \subset \overline{X_0X_1}$: We draw the curve equidistant from g_x and

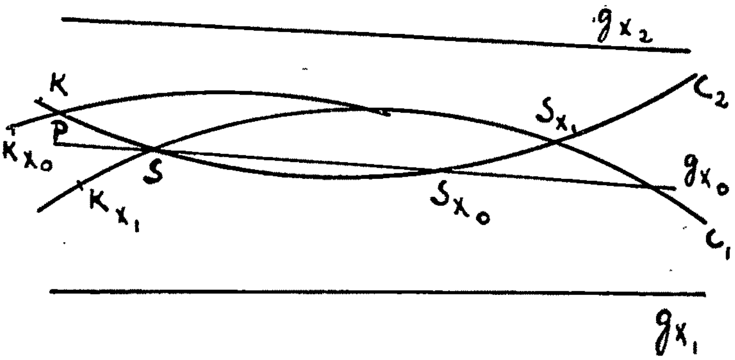


Fig. 4.

through S which may intersect c_2 in S_x (in addition to S). The motion of Σ composed of the translation along g_{x_2} carrying S into S_x and the translation along g_x carrying S_x into S may move K into K_x. The points K_x are on the circle $K(S, K)$; K_x depends in a continuous manner on g_x and does not remain fixed if X varies from X_1 to X_0. For, if $g_x = g_{x_0}$ ($= c_{x_0}$), the point S_{x_0} is on c_2 between S and S_{x_1}; therefore, the translation along g_{x_2} moving S into S_{x_0} moves K into a point P on c_2 between K and S, and $(S_{x_0} \rightarrow S)$ moves P into K_{x_0} which, therefore, is between g_{x_0} and g_{x_2} (or on g_{x_0}, if g_{x_0} passes through S and S_{x_1}) and different from K_{x_1}.

Thus there exist rotations with S as center transforming K into all points of an arc of $K(S, K)$ leading from K_{x_1} to K_{x_0}; hence, the full group of rotations around S exists. Since, by suitable translations along g and h the point S can be moved into each point of the plane, all rotations around each point exist. It is well known that this implies that the metric is hyperbolic.[18]

---

[18] The only place where this fact is stated exactly in the form used here, seems to be [5], Anhang IV. Of course this is much too deep a result for our purpose; one verifies easily in a direct way that the axioms of the hyperbolic plane geometry hold (in any of the current forms).

The theorems (4. 1) and (4. 4) and the examples (2. 11) and (3. 5) give together

If there exist all translations along two shortest lines which are neither parallel nor asymptotes to each other, then the metric is Desarguesian, namely, either Minkowskian or hyperbolic. If h and g are asymptotes or parallel to each other the metric is not necessarily Desarguesian, even if one assumes in addition that the parallel axiom either in the Euclidean or in the hyperbolic form holds with respect to each shortest line.

INSTITUTE FOR ADVANCED STUDY.

REFERENCES.

[1] Bonnesen, T. und W. Fenchel, "Theorie der konvexen Körper," *Ergebnisse der Mathematik*, vol. 3 (1934), fasc. 1.

[2] Busemann, H., "Pasch'sches Axiom und Zweidimensionalität," *Mathematische Annalen*, Bd. 107 (1932), pp. 324-328.

[3] Busemann, H., "Über die Geometrien, in denen die 'Kreise mit unendlichem Radius' die kürzesten Linien sind," *Mathematische Annalen*, Bd. 106 (1932), pp. 140-160.

[4] Funk, P., "Über die Geometrien, bei denen die Geraden die kürzesten Linien sind und die Äquidististanten zu einer Geraden wieder Gerade sind," *Monatshefte für Mathematik und Physik*, Bd. 37 (1930), pp. 153-158.

[5] Hilbert, D., *Grundlagen der Geometrie*, 7th edition, Leipzig, 1930.

[6] Liebmann, H., *Nichteuklidische Geometrie*, 3rd edition, Berlin, 1923.

[7] Menger, K., "Untersuchungen über allgemeine Metrik," *Mathematische Annalen*, Bd. 100 (1928), pp. 75-163.

[8] Minkowski, H., *Geometrie der Zahlen*, Leipzig, 1910.

[9] Reinhardt, K., "Über einen Satz von Herrn H. Tietze," *Jahresber. Deutsch. Mathem. Ver.*, Bd. 38 (1929), pp. 191-192.

[10] Rosenthal, A., "Die Translationsordnung ebener Kurven," *Monatshefte für Mathematik und Physik*, Bd. 45 (1936), pp. 76-91.

[11] Schilling, F., *Projektive und nichteuklidische Geometrie*, Bd. 1, Leipzig, 1931.

[12] Tietze, H., "Eine charakteristische Eigenschaft der abgeschlossenen konvexen Punktmengen," *Mathematische Annalen*, Bd. 99 (1928), pp. 394-398.

[13] Veblen, O. and I. W. Young, *Projective Geometry*, vol. II, Boston, 1918.

CALCUTTA UNIVERSITY Sec. No. Book No.

AN INTEGRO-DIFFERENTIAL BOUNDARY VALUE PROBLEM.*

By WILLIAM T. REID.

1. Introduction.

Lichtenstein [12][1] has treated by means of the theory of quadratic forms in infinitely many variables a boundary value problem involving a single integro-differential equation of the second order and a special set of two-point boundary conditions. Under certain conditions he proved the existence of infinitely many characteristic values, and also established an expansion theorem for functions in terms of the corresponding characteristic solutions. More recently, Lichtenstein [13] has used the results of his previous paper to prove by expansion methods sufficient conditions for a weak relative minimum in the isoperimetric problem of the calculus of variations. Courant ([6], Sections 5 and 13) has treated by the method of difference equations an integro-differential boundary problem similar to that considered by Lichtenstein.

Tamarkin [21][2] has developed in two papers a somewhat general theory for an integro-differential boundary problem consisting of a single linear integro-differential equation of the n-th order and boundary conditions which involve not only the end-values of the solution and its first $n-1$ derivatives at two points, but also involve integral terms containing the solution functions. The outstanding feature of Tamarkin's treatment is the repeated use of the notion of a Green's function for the integro-differential system. In an unpublished dissertation Jonah [11] has proved the existence of a Green's matrix for a boundary problem associated with a system of integro-differential equations of the first order, and has extended the principal results of Tamarkin's papers to such a system.

The boundary problem which is here considered may be formulated as follows. Let

$$(1.1)\qquad I[\eta]=2Q[\eta(a),\eta(b)]+\int_a^b 2\omega(x,\eta,\eta')\,dx + \int_a^b\int_a^b M_{ij}(x,t)\eta_i(x)\eta_j(t)\,dx\,dt,$$

---

* The present paper contains the revised form of results presented to the American Mathematical Society under the titles "*An integro-differential boundary problem,*" December 28, 1934 and "*A boundary value problem associated with the calculus of variations,* II," April 19, 1935. Received by the Editors, September 29, 1937.

[1] Numerals in square brackets refer to the bibliography at the end of the present paper.

[2] Other references to literature on integro-differential boundary problems will be found in the introduction of Tamarkin's first paper.

where $\eta \equiv [\eta_i(x)]$ $(i = 1, \cdots, n)$ and $\eta_i(x)$ are real-single-valued functions on $a \leq x \leq b$, ω and Q are quadratic forms in the $2n$ variables η_i, η'_i and $\eta_i(a)$, $\eta_i(b)$, respectively. This paper treats the system consisting of the Euler-Lagrange integro-differential equations and transversality conditions for the problem of minimizing $I[\eta]$ in a class of arcs

$$(1.2) \qquad \eta_i = \eta_i(x) \qquad (i = 1, \cdots, n;\ a \leq x \leq b)$$

which satisfy a set of ordinary linear differential equations of the first order

$$(1.3) \qquad \Phi_\alpha(x, \eta, \eta') \equiv \Phi_{\alpha\pi_i}(x)\eta'_i + \Phi_{\alpha\eta_i}(x)\eta_i = 0 \qquad (\alpha = 1, \cdots, m < n),$$

together with linear homogeneous end-conditions

$$(1.4)\ \Psi_\gamma[\eta(a), \eta(b)] \equiv \Psi_{\gamma:ia}\eta_i(a) + \Psi_{\gamma;ib}\eta_i(b) = 0\ (\gamma = 1, \cdots, p \leq 2n),$$

and which are such that

$$(1.5) \qquad K[\eta] \equiv \int_a^b (\eta_1^2 + \cdots + \eta_n^2)\,dx = 1.$$

In the particular case when $M_{ij}(x, t) \equiv 0$ the expression (1.1) is of the form of the second variation of a problem of Bolza in the calculus of variations and the above described boundary problem is the so-called accessory boundary problem, and has been treated by various authors (see Morse [15] and [16], Reid [18] and [19], Hu [9], Hölder [8], Birkhoff and Hestenes [1], and Wiggin [22]). In the general case, expression (1.1) is the form of the second variation of a more general problem of the calculus of variations. The boundary value problem described above includes as a very special case the problem considered by Lichtenstein. It also includes a class of problems associated with a single self-adjoint linear integro-differential equation of even order. It is to be noted, however, that the boundary conditions of this problem are two-point conditions, and hence are not of as general a character as those of the problem treated by Tamarkin.

The hypotheses upon which the analysis is based are stated in Section 2 and in Section 3 some properties of the boundary problem are discussed. In Section 4 there are defined successive classes S_s $(s = 1, 2, \cdots)$ of arcs η in which we consider the problem of minimizing $I[\eta]$, and it is shown that the greatest lower bound of $I[\eta]$ in S_s is a characteristic value of our problem. The method of proof is similar to that previously used by the author [18] for the differential problem to which the above problem reduces whenever

$M_{ij}(x,t) \equiv 0$. Use is now made of a Green's matrix for an integro-differential system as introduced by Tamarkin and Jonah.

Sections 5 and 6 are devoted, respectively, to comparison and oscillation theorems. These theorems are generalizations of the Sturmian comparison and oscillation theorems for a single second-order linear differential equation (see, for example, Ince [10], Chapter X). These theorems have been previously established for a problem of the above sort involving differential equations, that is, for which $M_{ij}(x,t) \equiv 0$, and which satisfies certain additional normality assumptions.[3] *The present paper gives for the first time such theorems not involving any assumptions of normality on sub-intervals. The definitions of focal and conjugate points given in Section* 6 *are also new.* It is to be remarked that for an integro-differential system of the sort here considered, even though it be identically normal, one can not in general define the focal points as zeros of a certain determinant. In fact, so far as the author knows, comparison and oscillation theorems have not been previously given for even the simplest type of integro-differential system here considered which does not reduce to a differential system, that is, for which $M_{ij}(x,t) \not\equiv 0$.

In Section 7 we consider an integro-differential problem which is in general non-linear in the parameter. There are obtained theorems on the existence of real characteristic values, as well as comparison and oscillation theorems, again without any assumption of normality on sub-intervals. Of extreme significance is the method of proof used. There is associated with the given problem an auxiliary problem which is linear in a second characteristic parameter, and the characteristic values of the new problem are considered as functions of the original parameter. Comparison and oscillation theorems are immediate consequences of the corresponding theorems for the associated problem and, as proved in Sections 5 and 6, for this latter problem these theorems follow from the extremizing properties of the characteristic values. This method of making the theory of a problem non-linear in the parameter depend upon the corresponding theory for a problem linear in a second parameter seems to be highly significant, in view of the fact that problems which are linear in the parameter lend themselves readily to treatment by diverse methods.[4] As far as the author knows, however, the method has not been used previously for even the simplest problem of the type here treated, even in the case of a self-adjoint differential equation of the second order with separated end-conditions which satisfies the conditions prescribed in the Sturmian theory.

---

[3] See Morse [14], [15], [16], and Hu [9]. Morse [16] has stated his results for a differential problem which involves no auxiliary differential equations $\Phi_a = 0$, and which satisfies the hypothesis of Theorem 7 of the present paper. As he points out, if suitable normality assumptions are made, his methods extend immediately to the more general differential boundary value problem. Birkhoff and Hestenes [1] have obtained these results for a differential system involving no auxiliary differential equations $\Phi_a = 0$.

[4] By a similar treatment one may readily extend some of the important results of

Finally, in Section 8 there is discussed briefly the idea of a Green's matrix for a system of integro-differential equations of the first order together with a set of two-point boundary conditions, and the fundamental theorems concerning a problem and its adjoint problem are given.

2. Notation and preliminary remarks. Throughout the first seven sections of this paper the following subscripts have the ranges indicated: $i, j, k = 1, \cdots, n$; $\alpha, \beta = 1, \cdots, m$; $\sigma, \tau = 1, \cdots, 2n$; $\gamma, \nu = 1, \cdots, p$; $\theta, \phi = 1, \cdots, 2n - p$. The repetition of a subscript in a single term of an expression will denote summation with respect to that subscript over its range of definition. Partial derivatives of $\omega(x, \eta, \pi)$, $\Phi_\alpha(x, \eta, \pi)$ with respect to the variables η_i, π_i will be denoted by writing these variables as subscripts; correspondingly, derivatives of Q and Ψ_γ with respect to the arguments $\eta_i(a)$, $\eta_i(b)$ will be denoted by Q_{ia}, $\Psi_{\gamma;ia}$, Q_{ib}, $\Psi_{\gamma;ib}$, respectively.

The analysis of the paper is based upon the following hypotheses:

(H_1) The coefficients of the quadratic form $\omega(x, \eta, \pi)$ and the linear expressions $\Phi_\alpha(x, \eta, \pi)$ are real-single-valued functions of x on ab. The functions $\omega_{\pi_i \pi_j}$, $\omega_{\pi_i \eta_j}$, $\Phi_{\alpha \pi_j}$ are of class C^1, and the functions $\omega_{\eta_i \eta_j}$, $\Phi_{\alpha \eta_j}$ are continuous on this interval. The functions $M_{ij}(x, t)$ are continuous on $a \leq x$, $t \leq b$, and $M_{ij}(x, t) = M_{ji}(t, x)$. Finally, the matrix $\| \Phi_{\alpha \pi_j} \|$ is of rank m on ab, the coefficients of the quadratic form Q and the linear homogeneous expressions Ψ_γ are real constants, and the matrix $\| \Psi_{\gamma;ia} \Psi_{\gamma;ib} \|$ has rank p.

An arc $\eta \equiv [\eta_i(x)]$ will be called *differentially admissible* if the functions $\eta_i(x)$ are of class D^1 on ab, and satisfy the equations $\Phi_\alpha = 0$ on this interval. An arc whose end values at a and b satisfy $\Psi_\gamma = 0$ will be said to be *terminally admissible.* Finally, an arc which is both differentially and terminally admissible will be called *admissible.*

(H_2) The quadratic form $\omega_{\pi_i \pi_j}(x) u_i u_j$ is positive for values $(u_i) \neq (0_i)$ and satisfying $\Phi_{\alpha \pi_j}(x) u_j = 0$ $(\alpha = 1, \cdots, m)$.

This hypothesis implies, in particular, that the matrix

$$(2.1) \qquad \left\| \begin{matrix} \omega_{\pi_i \pi_j} & \Phi_{\alpha \pi_i} \\ \Phi_{\beta r_j} & O_{\beta \alpha} \end{matrix} \right\| \qquad (i, j = 1, \cdots, n;\ \alpha, \beta = 1, \cdots, m)$$

is non-singular on ab.

---

the Hilbert-Schmidt theory of linear equations to corresponding integral equations involving the parameter non-linearly.

(H_3) There exist p differentially admissible arcs $\eta_i = \eta_{i\nu}(x)$ $(\nu = 1, \cdots, p)$ such that the determinant $| \Psi_\gamma[\eta_\nu(a), \eta_\nu(b)] |$ is different from zero.

Hypothesis (H_3) is a condition of normality with respect to the differential equations (1. 3) and the end-conditions (1. 4). It implies, in particular, that the conditions Ψ_γ are linearly independent. One may show that if η is a minimizing arc for the problem of the calculus of variations defined in Section 1, and satisfying the above hypotheses, then there exists a constant λ and functions $\mu_\alpha(x)$ such that if we define

$$\Omega(x, \eta, \pi, \mu) = \omega(x, \eta, \pi) + \mu_\alpha \Phi_\alpha(x, \eta, \pi), \tag{2. 2}$$

$$J_i(\eta, \mu) = (d/dx)\Omega_{\pi_i}(x, \eta, \eta', \mu) - \Omega_{\eta_i}(x, \eta, \eta', \mu) - \int_a^b M_{ij}(x, t)\eta_j(t)\,dt, \tag{2. 3}$$

then the set η, μ, λ satisfies the integro-differential system

$$J_i(\eta, \mu) + \lambda\eta_i = 0, \qquad \Phi_\alpha(x, \eta, \eta') = 0; \tag{2. 4}$$

moreover, there exist constants d_γ satisfying with the end values of η_i, μ_α the end conditions

$$\begin{aligned} Q_{ia}[\eta] + d_\gamma \Psi_{\gamma;ia} - \Omega_{\pi_i}(x, \eta, \eta', \mu)\,|^{x=a} &= 0, \\ Q_{ib}[\eta] + d_\gamma \Psi_{\gamma;ib} + {}_{ib} + \Omega_{\pi_i}(x, \eta, \eta', \mu)\,|^{x=b} &= 0, \\ \Psi_\gamma[\eta(a), \eta(b)] &= 0. \end{aligned} \tag{2. 5}$$

Since the matrix (2. 1) is non-singular, the set of $m + n$ equations

$$\zeta_i = \Omega_{\pi_i}(x, \eta, \pi, \mu), \ \Phi_\alpha(x, \eta, \pi) = 0 \quad (\alpha = 1, \cdots, m; \ i = 1, \cdots, n) \tag{2. 6}$$

has unique solutions

$$\pi_i = A_{ij}(x)\eta_j + B_{ij}(x)\zeta_j, \qquad \mu_\alpha = l_{\alpha j}(x)\eta_j + m_{\alpha j}(x)\zeta_j. \tag{2. 7}$$

When these values are substituted in $\Omega_{\eta_i}(x, \eta, \pi, \mu)$ we obtain

$$\Omega_{\eta_i}(x, \eta, \pi, \mu) = C_{ij}(x)\eta_j - A_{ji}(x)\zeta_j. \tag{2. 8}$$

In view of (H_1) and (H_2) the functions A_{ij}, B_{ij}, C_{ij} are of class C^1 on ab; moreover, the matrices $\| B_{ij} \|$ and $\| C_{ij} \|$ are symmetric and $\| B_{ij} \|$ is of rank $n - m$ on ab.

The system (2. 4) is therefore equivalent to the system

$$\begin{aligned} L_i[\eta, \zeta] &\equiv \eta'_i - A_{ij}(x)\eta_j - B_{ij}(x)\zeta_j = 0, \\ L_{n+i}[\eta, \zeta|\lambda] &\equiv \zeta'_i - C_{ij}(x)\eta_j + A_{ji}(x)\zeta_j \\ &\quad - \int_a^b M_{ij}(x, t)\eta_j(t)\,dt + \lambda\eta_i = 0. \end{aligned} \tag{2. 4'}$$

Now if $c_i = c_{i\theta}$, $d_i = d_{i\theta}$ $(\theta = 1,\cdots,2n-p)$ are linearly independent solutions of the equations

$$\Psi_{\gamma;ia}c_i + \Psi_{\gamma;ib}d_i = 0 \qquad (\gamma = 1,\cdots,p),$$

the boundary conditions (2. 5) are equivalent to the linearly independent set

$$\begin{aligned} &s_\gamma[\eta,\zeta] \equiv \Psi_\gamma[\eta(a),\eta(b)] = 0, \\ &s_{p+\theta}[\eta,\zeta] \equiv c_{i\theta}\{Q_{ia}[\eta] - \zeta_i(a)\} + d_{i\theta}\{Q_{ib}[\eta] + \zeta_i(b)\} = 0. \end{aligned} \tag{2. 5'}$$

For brevity, we shall speak of the boundary value problem (2. 4), (2. 5), or the equivalent problem (2. 4′), (2. 5′), as the boundary value problem B. Corresponding to these two forms of B, the term characteristic solution will be used to denote a set η_i, μ_a or a set η_i, ζ_i, where in each case the functions of the set are not all identically zero on ab, and the set satisfies for a corresponding value λ the integro-differential equations and two-point boundary conditions of B. If λ is a value for which there exist q linearly independent characteristic solutions, this value will be called a characteristic value of B of index q.

The above assumption (H_3) is equivalent to the assumption that there is no characteristic solution η_i, ζ_i of B for which $\eta_i \equiv 0$ on ab (see Bliss [3], p. 693; [4], p. 48). As is customary (see Reid [20], p. 575), we shall say that the order of anormality on the interval ab of the integro-differential equations of B is equal to r if on this interval there are exactly r linearly independent solutions $\eta_i \equiv 0$, $\zeta_i = v_{ih}(x)$ $(h = 1,\cdots,r)$ of (2. 4′). It is to be noted that the value of r is independent of λ. It follows readily that $r \leq m$; moreover, if η is an arbitrary differentially admissible arc and x_1, x_2 are any two points on ab, then

$$v_{ih}(x)\eta_i(x)\Big|_{x_1}^{x_2} = 0 \qquad (h = 1,\cdots,r). \tag{2. 9}$$

As a consequence of these relations, we have that for a system of the above type which satisfies (H_1) and (H_2) the additional hypothesis (H_3) is equivalent to the assumption that the matrix

$$\left\| \begin{matrix} \Psi_{\gamma;ja} & \Psi_{\gamma;jb} \\ -v_{jh}(a) & v_{jh}(b) \end{matrix} \right\| \qquad (j = 1,\cdots,n;\ h = 1,\cdots,r;\ \gamma = 1,\cdots,p) \tag{2. 10}$$

has rank $p + r$. This implies, in particular, that $p \leq 2n - r$.

Now the class of admissible arcs η for a problem B is unchanged when the conditions $\Psi_\gamma = 0$ are replaced by conditions of the form

$$\Psi_\gamma + e_{\gamma h}[-v_{jh}(a)\eta_j(a) + v_{jh}(b)\eta_j(b)] = 0 \qquad (\gamma = 1,\cdots,p), \tag{2. 11}$$

where $e_{\gamma h}$ $(\gamma = 1, \cdots, p;\ h = 1, \cdots, r)$ are arbitrary constants. If η_i, ζ_i is a characteristic solution of the original problem B, there are constants e_h such that $\eta_i, \zeta_i + v_{ih}e_h$ is a characteristic solution of the modified problem B involving the end-conditions (2.11). Consequently, we shall not distinguish between two boundary value problems B_1 and B_2 which satisfy the above hypotheses and differ only in the end-conditions $\Psi_\gamma{}^1 = 0$ and $\Psi_\gamma{}^2 = 0$ $(\gamma = 1, \cdots, p)$, respectively, and these conditions are such that the matrix

$$(2.12) \quad \left\| \begin{matrix} \Psi^1{}_{\gamma;ja} & \Psi^1{}_{\gamma;jb} \\ \Psi^2{}_{\gamma;ja} & \Psi^2{}_{\gamma;jb} \\ -v_{jh}(a) & v_{jh}(b) \end{matrix} \right\| \quad (j = 1, \cdots, n;\ \gamma = 1, \cdots, p;\ h = 1, \cdots, r)$$

has rank $p + r$.

If a problem B satisfies (H_1) and (H_2), but not (H_3), then the matrix (2.10) has rank $p + r - k$, where $0 < k \leq p$. By deleting k of the end-conditions $\Psi_\gamma = 0$ $(\gamma = 1, \cdots, p)$ one may then obtain a problem B^1 which satisfies hypothesis (H_3), and which is equivalent to B in the sense that an arc η is admissible for B^1 if and only if it is admissible for B. Such a problem B^1 will be called the *normal boundary problem determined by the end-conditions* $\Psi_\gamma = 0$ $(\gamma = 1, \cdots, p)$.

For brevity, we shall refer to hypotheses (H_1), (H_2) and (H_3) as simply hypotheses (H).

3. Properties of the boundary value problem *B*. For brevity we shall set

$$M[u; v] = M[v; u] = \int_a^b \int_a^b u_i(x) M_{ij}(x, t) v_j(t)\, dx dt;$$

$$K[u; v] = K[v; u] = \int_a^b u_i(x) v_i(x)\, dx$$

$$Q[u; v] = Q[v; u] = (1/2)\,(u_i(a) Q_{ia}[v] + u_i(b) Q_{ib}[v]);$$

$$\begin{aligned} I[u; v] &= I[v; u] \\ &= Q[u; v] + \int_a^b [u'_i \omega_{\pi_i}(x, v, v') + u_i \omega_{\eta_i}(x, v, v')]\, dx + M[u; v]; \end{aligned}$$

$$M[u] = M[u; u]; \quad K[u] = K[u; u]; \quad Q[u] = Q[u; u]; \quad I[u] = I[u; u];$$

$$I[u; v|\lambda] = I[u; v] - \lambda K[u; v], \quad I[u|\lambda] = I[u] - \lambda K[u].$$

In view of (2.6) and (2.7) we have:

LEMMA 3.1. *If η is a differentially admissible arc and μ_a an arbitrary set of functions of class D, the functions ζ_i defined by* (2.6) *are such that*

$$(3.1) \qquad \int_a^b 2\omega(x, \eta, \eta')\, dx = \int_a^b [B_{ij}(x) \zeta_i \zeta_j + C_{ij}(x) \eta_i \eta_j]\, dx.$$

The following properties of a boundary problem B satisfying hypotheses (H) will be stated without proof, since they follow as for the differential problem to which B reduces when $M_{ij}(x,t) \equiv 0$ (see, in particular, Hu [9] and Reid [18]).

THEOREM 3.1. *If η_i, ζ_i is a solution of the non-homogeneous system*

$$(3.2) \qquad L_i[\eta,\zeta] = 0, \quad L_{n+i}[\eta,\zeta|\lambda] + f_i(x) = 0, \quad s_\sigma[\eta,\zeta] = 0,$$

and u_i is an arbitrary differentially admissible arc, then $I[\eta;u|\lambda] - K[f;u] = 0$.

COROLLARY 1. *If η_i, ζ_i is a solution of (3.2), then $I[\eta|\lambda] - K[f;\eta] = 0$.*

COROLLARY 2. *If η_i, ζ_i is a characteristic solution of B corresponding to a characteristic value λ, then $I[\eta|\lambda] = 0$.*

THEOREM 3.2. *If η_i, ζ_i and η^*_i, ζ^*_i are solutions of system (3.2) for sets of functions $f_i(x), f^*_i(x)$, respectively, then $K[\eta;f^*] - K[\eta^*;f] = 0$.*

COROLLARY. *If η_i, ζ_i and η^*_i, ζ^*_i are characteristic solutions of B for distinct characteristic values λ and λ^*, then $K[\eta;\eta^*] = 0$.*

THEOREM 3.3. *The boundary problem B has only real characteristic values, and the corresponding characteristic solutions may be chosen real.*

LEMMA 3.2. *There exist positive constants m_0, l_0 such that for arbitrary differentially admissible arcs η we have*

$$(3.3) \qquad I[\eta] \geq \int_a^b [m_0\eta'_i\eta'_i - l_0\eta_i\eta_i]\,dx.$$

Let c be a constant such that $|2Q[\eta_{ia},\eta_{ib}]| \leq c[\eta_{ia}\eta_{ia} + \eta_{ib}\eta_{ib}]$, and choose a function $\phi(x)$ of class C^1 such that $\phi(b) = -1$, $\phi(a) = 1$. Moreover denote by l_1 a value such that $|M[\eta]| \leq l_1 K[\eta]$ for arbitrary continuous functions $\eta_i(x)$. It follows from Schwarz' inequality that l_1 may be chosen as $M_0(b-a)$, where M_0 is such that $|M_{ij}(x,t)| \leq M_0/n$ $(i,j = 1,\cdots,n;$ $a \leq x, t \leq b)$. Then for differentially admissible arcs η we have

$$(3.4) \qquad I[\eta] \geq \int_a^b 2\omega^*(x,\eta,\eta')\,dx,$$

when $2\omega^* = 2\omega + c[\phi'(x)\eta_i\eta_i + 2\phi(x)\eta_i\eta'_i] - l_1\eta_i\eta_i$. For a given differentially admissible η and arbitrary multipliers μ_α of class D, let A^*_{ij}, B^*_{ij}, C^*_{ij} denote the coefficients in (2.7) and (2.8) when ω is replaced by ω^* in equations (2.6). It is seen that $B^*_{ij} = B_{ij}$. It is then a consequence of Lemma 3.1 and (3.4) that

$$I[\eta] \geq \int_a^b [B_{ij}(x)\zeta_i\zeta_j + C^*_{ij}(x)\eta_i\eta_j]\,dx.$$

For arbitrary functions ζ_i the quantities $p_i = B_{ij}\zeta_j$ are such that $p_i\omega_{\pi_i\pi_j}p_j = B_{ij}\zeta_i\zeta_j$, and $\Phi_{\alpha\pi_j}p_j = 0$ $(\alpha = 1, \cdots, m)$. In view of (H_2) there is a positive constant l such that $B_{ij}\zeta_i\zeta_j \geq l p_i p_i$. Now $p_i = B_{ij}\zeta_j = \eta'_i - A^*_{ij}\eta_j$, and therefore $p_i p_i \geq (1/2)\eta'_i\eta'_i - A^*_{ki}A^*_{kj}\eta_i\eta_j$.[5] Inequality (3.3) is then seen to be valid if $m_0 = l/2$, and l_0 is a constant such that for $a \leq x \leq b$ the relation $|[C^*_{ij} - lA^*_{ki}A^*_{kj}]u_iu_j| \leq l_0u_iu_i$ holds for arbitrary sets u_i.

COROLLARY. *There exists a constant* λ_0 *such that* $I[\eta|\lambda_0] > 0$ *for arbitrary non-identically vanishing differenticlly admissible arcs* η.

LEMMA 3.3. *If* η_i, ζ_i *is a characteristic solution of* B *for a value* λ *which is normed so that* $K[\eta] = 1$, *then*

$$(3.5) \quad \eta_i(x)\eta_i(x) \leq 2/(b-a) + 2(\lambda + l_0)(b-a)/m_0, \quad a \leq x \leq b,$$

where m_0, l_0 *are the constants of Lemma* 3.2.

If x_1, x are two points of ab, by Schwarz' inequality we have

$$[\eta_i(x) - \eta_i(x_1)]^2 = [\int_{x_1}^{x} \eta'_i(t)\,dt]^2 \leq |x - x_1| \; |\int_{x_1}^{x} \eta'^2_i dt|$$
$$\leq (b-a)\int_a^b \eta'^2_i dt, \quad (i = 1, \cdots, n),$$

and hence, by Lemma 3.2,

$$\sum_{i=1}^{n} [\eta_i(x) - \eta_i(x_1)]^2 \leq (b-a)\int_a^b \eta'_i\eta'_i dt \leq [(b-a)/m_0](I[\eta] + l_0K[\eta])$$
$$\leq [(b-a)/m_0](\lambda + l_0).$$

Since $K[\eta] = 1$ we may choose x_1 such that $\eta_i(x_1)\eta_i(x_1) \leq 1/(b-a)$. Relation (3.5) is then a ready consequence of the elementary inequality

$$\eta_i(x)\eta_i(x) \leq 2\{\eta_i(x_1)\eta_i(x_1) + \sum_{i=1}^{n} [\eta_i(x) - \eta_i(x_1)]^2\}.$$

The remainder of the present section depends upon the results of Section 8. According to the definition of that section, the integro-differential system B may be shown to be self-adjoint with respect to the constant transformation matrix

---

[5] This relation is an immediate consequence of the elementary inequality

$$a^2 \leq 2[(a-b)^2 + b^2],$$

which is satisfied by arbitrary real constants a and b.

$$(3.6)\qquad \| T_{\sigma\tau} \| = \begin{Vmatrix} 0_{ij} & \delta_{ij} \\ -\delta_{ij} & 0_{ij} \end{Vmatrix} \qquad (i, j = 1, \cdots, n;\ \sigma, \tau = 1, \cdots, 2n).$$

If λ is not a characteristic value of B, it follows that the non-homogeneous system

$$(3.7)\qquad L_i[\eta, \zeta] = 0,\ L_{n+i}[\eta, \zeta | \lambda] = k_i(x), \qquad s_\sigma[\eta, \zeta] = 0$$

has for given continuous functions $k_i(x)$ a unique solution η_i, ζ_i. Moreover, this solution is given by

$$(3.8)\quad \eta_i(x|\lambda) = \int_a^b G^2{}_{ij}(x, t|\lambda) k_j(t)\, dt,\ \ \zeta_i(x|\lambda) = \int_a^b G^4{}_{ij}(x, t|\lambda) k_j(t)\, dt,$$

where

$$(3.9)\qquad \| G_{\sigma\tau}(x, t|\lambda) \| = \begin{Vmatrix} G^1{}_{ij}(x, t|\lambda) & G^2{}_{ij}(x, t|\lambda) \\ G^3{}_{ij}(x, t|\lambda) & G^4{}_{ij}(x, t|\lambda) \end{Vmatrix}$$

is the Green's matrix of the system (2. 4′), (2. 5′) for this value of λ.

4. Existence and properties of characteristic values. There will now be defined a sequence of classes of arcs, and we shall consider the problem of minimizing $I[\eta]$ in each of these classes. The class S_1 is defined as the totality of admissible arcs which satisfy the relation $K[\eta] = 1$. In view of (H_3) the class S_1 is non-vacuous (see Bliss [3], p. 694). We shall first prove the following result

THEOREM 4. 1. *Suppose the boundary problem B satisfies hypotheses (H), and denote by Λ_1 the greatest lower bound of $I[\eta]$ in the class S_1. Then $\lambda = \Lambda_1$ is the smallest characteristic value of B.*

In view of Lemma 3. 2 the greatest lower bound of $I[\eta]$ in S_1, Λ_1, is finite. It also follows from Corollary 2 to Theorem 3. 1 that there is no characteristic value of B less than Λ_1. It will now be proved by indirect argument that $\lambda = \Lambda_1$ is a characteristic value. Let $u_\nu = [u_{i\nu}]$ $(\nu = 1, 2, \cdots)$ be a sequence of arcs belonging to S_1 and such that $I[u_\nu] \to \Lambda_1$ as $\nu \to \infty$.

Now consider the non-homogeneous system

$$(4.1)\qquad L_i[\eta, \zeta] = 0, \qquad L_{n+i}[\eta, \zeta | \Lambda_1] = u_{i\nu}(x), \qquad s_\sigma[\eta, \zeta] = 0,$$

that is, the system (3. 7) for $\lambda = \Lambda_1$, $k_i = u_{i\nu}$. If $\lambda = \Lambda_1$ is not a characteristic value of B the system (4. 1) has a unique solution $\eta_i = U_{i\nu}(x)$, $\zeta_i = V_{i\nu}(x)$, $(\nu = 1, 2, \cdots)$, given by the corresponding relations (3. 8). It is a consequence of elementary inequalities that there exist constants l_1 and l_2 depending solely upon the value of Λ_1 and such that

$$(4.2) \qquad |U_{i\nu}(x)| \leq l_1 K[u_\nu] + l_2 = l_1 + l_2 \qquad (\nu = 1, 2, \cdots).$$

Now set

$$(4.3) \qquad w_{i\nu}(x) = u_{i\nu}(x) + cU_{i\nu}(x) \qquad (i = 1, \cdots, n;\ \nu = 1, 2, \cdots),$$

where c is an arbitrary real constant. It is readily verified with the aid of Theorem 3.1 that

$$(4.4) \qquad I[w_\nu|\Lambda_1] = I[u_\nu|\Lambda_1] - 2c - c^2 K[u_\nu; U_\nu].$$

By Schwarz' inequality we have $(K[u_\nu; U_\nu])^2 \leq K[u_\nu]K[U_\nu] = K[U_\nu]$, and it therefore follows by (4.2) that there is a constant $l_3 > 0$ such that $|K[u_\nu; U_\nu]| \leq l_3$ $(\nu = 1, 2, \cdots)$. Now $I[u_\nu|\Lambda_1] = I[u_\nu] - \Lambda_1$ tends to zero as $\nu \to \infty$. Hence if c be chosen so that $0 < c < 2/l_3$, it follows from (4.4) that $I[w_\nu|\Lambda_1] = I[w_\nu] - \Lambda_1 K[w_\nu]$ is negative for ν sufficiently large. This, however, contradicts the assumption that Λ_1 is the greatest lower bound of $I[\eta]$ in S_1. Hence $\lambda = \Lambda_1$ is a characteristic value of B, and Theorem 4.1 is proved.

Now suppose classes $S_1, \cdots, S_{s-1}$ $(s \geq 2)$ have been defined, each of these classes is non-vacuous, and that $\lambda = \Lambda_t$ $(t = 1, \cdots, s-1)$, where Λ_t is the greatest lower bound of $I[\eta]$ in the class S_t, is a characteristic value of B of index ρ_t. Let $\lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_\rho$ $(\rho = \rho_1 + \rho_2 + \cdots + \rho_{s-1})$ denote these characteristic values each repeated a number of times equal to its index, and choose $\eta_i = y_{i\kappa}(x)$, $\zeta_i = z_{i\kappa}(x)$ as a characteristic solution of B for $\lambda = \lambda_\kappa$ $(\kappa = 1, \cdots, \rho)$. These solutions will be supposed to be orthonormal in the sense that $K[y_\kappa; y_{\kappa'}] = \delta_{\kappa\kappa'}$ $(\kappa, \kappa' = 1, \cdots, \rho)$. The class S_s is then defined as the sub-class of S_{s-1} consisting of all admissible arcs η which satisfy the relations

$$(4.5) \qquad K[y_\kappa; \eta] = 0 \qquad (\kappa = 1, \cdots, \rho)$$

The class S_s is well defined and non-vacuous (see Reid [19], p. 847). We shall now prove the following induction theorem.

THEOREM 4.2. *Suppose the boundary problem B satisfies hypotheses (H), and denote by Λ_s the greatest lower bound of $I[\eta]$ in the class S_s. Then $\lambda = \Lambda_s$ is a characteristic value of B, and $\Lambda_s > \Lambda_{s-1}$.*

To prove this theorem we shall consider the auxiliary problem of minimizing the expression $\boldsymbol{I}[\boldsymbol{\eta}] = I[\eta|\Lambda_s]$ in the class $\boldsymbol{S}_1$ of arcs

$$\boldsymbol{\eta} = (\eta_1, \cdots, \eta_n, \eta_{r+1}, \cdots, {}_{n+\rho}),$$

where the functions $\eta_1, \cdots, \eta_{n+\rho}$ are of class D^1 on ab, satisfy the differential equations

$$(1.3^*)\qquad \Phi_{\alpha\pi_i}(x)\eta'_i + \Phi_{\alpha\eta_i}(x)\eta_i = 0, \qquad \eta'_{n+\kappa} - y_{i\kappa}(x)\eta_i = 0$$
$$(\alpha = 1, \cdots, m;\ \kappa = 1, \cdots, \rho;\ i = 1, \cdots, n),$$

the end conditions

$$(1.4^*)\qquad \Psi_{\gamma;\,ia}\eta_i(a) + \Psi_{\gamma;\,ib}\eta_i(b) = 0, \qquad \eta_{n+\kappa}(a) - \eta_{n+\kappa}(b) = 0$$
$$(\gamma = 1, \cdots, p;\ \kappa = 1, \cdots, \rho),$$

and the norming condition

$$(1.5)\qquad \int_a^b [\eta_i\eta_i + \eta_{n+\kappa}\eta_{n+\kappa}]dx \equiv \boldsymbol{K}[\boldsymbol{\eta}] = 1.$$

If $\boldsymbol{B}$ denote the corresponding boundary value problem involving

$$\boldsymbol{\eta} = (\eta_1, \cdots, \eta_{n+\rho}), \qquad \boldsymbol{\zeta} = (\zeta_1, \cdots, \zeta_{n+\rho})$$

it is readily seen that $\boldsymbol{B}$ satisfies the corresponding hypotheses (H). Consequently, if $\boldsymbol{\Lambda}_1$ is the greatest lower bound of $\boldsymbol{I}[\boldsymbol{\eta}]$ in the class $\boldsymbol{S}_1$, it follows from the above theorem that $\boldsymbol{\Lambda}_1$ is the smallest characteristic value of $\boldsymbol{B}$. Suppose η is an arbitrary admissible arc for B which satisfies relations (4. 5), and let $\boldsymbol{\eta}$ denote the corresponding admissible arc for $\boldsymbol{B}$. In view of the relation $I[\eta|\Lambda_s] = \boldsymbol{I}[\boldsymbol{\eta}] \geqq \boldsymbol{\Lambda}_1\boldsymbol{K}[\boldsymbol{\eta}] \geqq \boldsymbol{\Lambda}_1 K[\eta]$, and the definition of Λ_s, it follows that $\boldsymbol{\Lambda}_1 = 0$. Hence there exist functions $\eta_1, \cdots, \eta_{n+\rho}, \zeta_1, \cdots, \zeta_{n+\rho}$ not all zero and such that

$$(4.6)\qquad \begin{aligned} &L_i[\eta, \zeta] = 0, \quad \eta'_{n+\kappa} = y_{i\kappa}(x)\eta_i, \quad \zeta'_{n+\kappa} = 0, \\ &L_{n+i}[\eta, \zeta|\Lambda_s] - y_{i\kappa}(x)\zeta_{n+\kappa} = 0, \\ &s_\sigma[\eta, \zeta] = 0, \quad \eta_{n+\kappa}(a) - \eta_{n+\kappa}(b) = 0, \\ &\qquad (i = 1, \cdots, n;\ \sigma = 1, \cdots, 2n;\ \kappa = 1, \cdots, \rho). \end{aligned}$$

It follows from the differential equations and boundary conditions satisfied by the functions $y_{i\kappa}$, $z_{i\kappa}$, together with Theorem 3. 2, that the constants $\zeta_{n+\kappa}$ are all zero, and consequently the functions η_i, ζ_i are not all identically zero on ab. Hence $\lambda = \Lambda_s$ is a characteristic value of B corresponding to which there are characteristic solutions η_i, ζ_i which satisfy relations (4. 5). This last condition implies $\Lambda_s \neq \Lambda_{s-1}$, and therefore $\Lambda_s > \Lambda_{s-1}$. This completes the proof of Theorem 4. 2.

For a given boundary value problem satisfying hypotheses (H) we shall denote by $\{\lambda_s\}$ $(s = 1, 2, \cdots)$ the totality of characteristic values, each repeated a number of times equal to its index, and the entire set ordered so that

$$(4.7)\qquad \lambda_1 \leqq \lambda_2 \leqq \cdots \leqq \lambda_s \leqq \cdots.$$

It will also be supposed that $\eta_i = y_{is}(x)$, $\zeta_i = z_{is}(x)$ is a solution of B for

$\lambda = \lambda_s$ $(s = 1, 2, \cdots)$, and that these solutions are orthonormal in the sense that

$$K[y_s ; y_t] = \delta_{st} \qquad (s, t = 1, 2, \cdots). \tag{4.8}$$

Finally, if we denote by $\mathfrak{S}_s$ the totality of admissible arcs η which satisfy the relations $K[\eta] = 1$, $K[y_\kappa ; \eta] = 0$ $(\kappa = 1, \cdots, s - 1)$, it is a consequence of the above theorems that λ_s is the minimum of $I[\eta]$ in the class $\mathfrak{S}_s$.

The following theorem is essentially the so-called maximum-minimum property of the characteristic values of B (see Courant-Hilbert [7], Chapter VI).

THEOREM 4. 3. *If s is a given positive integer and d_κ $(\kappa = 1, \cdots, s)$ are arbitrary real constants such that $d_1{}^2 + \cdots + d_s{}^2 = 1$, then the admissible arc $\eta_i = y_{i1}(x) d_1 + \cdots + y_{is}(x) d_s$ satisfies the relation $I[\eta] \leq \lambda_s$.*

For such an arc η it follows with the use of Theorem 3. 1 that

$$I[\eta] = \sum_{\kappa=1}^{s} \lambda_\kappa d_\kappa{}^2 \leq \lambda_s \sum_{\kappa=1}^{s} d_\kappa{}^2 = \lambda_s. \tag{4.9}$$

THEOREM 4. 4. *Suppose u is an admissible arc such that $K[u] = 1$, and that there is a positive integer s such that $I[u] = \lambda_s$, $K[y_\kappa ; u] = 0$ $(\kappa = 1, \cdots, s - 1)$. Then the functions $u_i(x)$ are of class C^1 and there exist functions $v_i(x)$ of class C^1 such that $\eta_i = u_i$, $\zeta_i = v_i$ is a characteristic solution of B for $\lambda = \lambda_s$.*

Let N be the greatest positive integer such that $\lambda_N = \lambda_s$. For

$$\eta_i = u_i - \sum_{\kappa=1}^{N} y_{i\kappa}(x) K[y_\kappa ; u] = u_i - \sum_{\kappa=s}^{N} y_{i\kappa}(x) K[y_\kappa ; u]$$

we have $K[y_\kappa ; u] = 0$ $(\kappa = 1, \cdots, N)$. Now suppose that the functions η_i are not all identically zero on ab. Then

$$0 < K[\eta] = 1 - \sum_{\kappa=s}^{N} (K[y_\kappa ; u])^2.$$

It is then a consequence of the minimizing property of λ_{N+1} that

$$I[\eta] \geq \lambda_{N+1} K[\eta] = \lambda_{N+1}\{1 - \sum_{\kappa=s}^{N} (K[y_\kappa ; u])^2\}. \tag{4.10}$$

On the other hand, since $y_{i\kappa}$, $z_{i\kappa}$ $(\kappa = s, \cdots, N)$ is a solution of B for $\lambda = \lambda_\kappa = \lambda_s$, we have by the use of Theorem 3. 1 that

$$I[\eta] = I[u] - \lambda_s \sum_{\kappa=s}^{N} (K[y_\kappa ; u])^2 = \lambda_s\{1 - \sum_{\kappa=s}^{N} (K[y_\kappa ; u])^2\}. \tag{4.11}$$

Relations (4.10) and (4.11) are contradictory since $\lambda_{N+1} > \lambda_s$. Hence $\eta_i(x) \equiv 0$, and the set $u_i,\ v_i = \sum_{\kappa=s}^{N} z_{i\kappa}(x) K[y_\kappa; u]$ satisfies the conclusion of the above theorem.

In closing this section, we shall indicate by simple examples that certain important properties of differential systems are not in general satisfied by the integro-differential system (2.4′), (2.5′). Firstly, it is in general not true that for a given λ there exists a solution η_i, ζ_i of the equations (2.4′) assuming prescribed initial values at $x = a$. This fact is illustrated by the example: $n = 1$, $Q \equiv 0$, $2\omega \equiv \eta'^2$, $M(x, t) = x + t$, $\lambda = 0$, $a = 0$, and b the positive root of the equation $2880 - 480b^4 - b^8 = 0$. We have $\eta' = \zeta$, and (2.4′) is equivalent to $\eta'' - \int_0^b (x + t)\eta(t)\, dt = 0$. It follows readily that any solution of this equation must be of the form $\alpha + \beta x + \gamma x^2 + \delta x^3$, and upon substitution it is found that for b a root of the above equation there exists a solution η if and only if α and β satisfy a linear relation. Moreover, $\eta = 4b^5 x^2 + (40 - 5b^4)x^3$ is a solution such that $\eta(0) = 0 = \eta'(0)$.

Secondly, consider the example $n = 1$, $Q \equiv 0$, $2\omega \equiv \eta'^2 - \eta^2$, $M(x, t) = 1/4\pi$, $\lambda = 0$, $a = 0$, $b = 4\pi$, $\Psi_1 \equiv \eta(a)$, $\Psi_2 \equiv \eta(b)$. The system (2.4′), (2.5′) is then equivalent to $\eta'' + \eta - (1/4\pi)\int_0^{4\pi} \eta(t)\, dt = 0$, $\eta(0) = 0 = \eta(4\pi)$. The integro-differential equation has the non-vanishing solution $\eta \equiv 1$, yet the corresponding functional $I[\eta]$ is negative for certain η's satisfying the end-conditions, for example, $I[\sin x/2] = -3\pi/2$.

It is to be noted, however, that in case $M_{ij}(x, t)$ is of product form $M_{ij}(x, t) \equiv \phi_{i1}(x)\phi_{j1}(t) + \cdots + \phi_{ig}(x)\phi_{jg}(t)$ the general integro-differential system (2.4′), (2.5′) may be reduced to an equivalent differential problem by the introduction of g additional variables. Let

$$\eta_{n+s}(x) = \int_a^x \phi_{is}(t)\eta_i(t)\, dt, \qquad (s = 1, \cdots, g).$$

Then the functional (1.1) may be written as

$$(4.12) \qquad 2Q[\eta(a), \eta(b)] + \sum_{s=1}^{g} [\eta_{n+s}(b)]^2 + \int_a^b 2\omega(x, \eta, \eta')\, dx,$$

where the symbol η still denotes the original set $\eta_1, \cdots, \eta_n$. Now consider the resulting differential problem consisting of the Euler equations and transversality conditions associated with (4.12) and the conditions (1.3), (1.4), (1.5) on the functions $\eta_1, \cdots, \eta_n$, together with the additional conditions $\Phi_{m+s} \equiv \eta'_{n+s} - \phi_{is}(x)\eta_i(x) = 0$, $\Psi_{p+s} \equiv \eta_{n+s}(a) = 0$, $(s = 1, \cdots, g)$. This system is equivalent to the original problem. It is to be noted however, that for this new problem the norming form (1.5) does not involve all of the variables $\eta_1, \cdots, \eta_{n+g}$ and hence the associated boundary problem is not of the form (2.4′), (2.5′); it is, however, of the form (7.7) considered in Section 7. The example of the preceding paragraph is of this type, and consequently we see why in general for the case of a single integro-differential equation of the second order with the boundary conditions $\eta(a) = 0 = \eta(b)$ the existence of a non-vanishing solution could not be expected to imply the non-negativeness of the associated functional $I[\eta]$.

5. Comparison theorems. A boundary problem B of the type treated in this paper depends upon differential forms $\omega(x,\eta,\eta')$, $\Phi_\alpha(x,\eta,\eta')$ $(\alpha = 1, \cdots, m)$, an end-form $Q[\eta(a),\eta(b)]$, end-relations $\Psi_\gamma[\eta(a),\eta(b)]$ $(\gamma = 1, \cdots, p)$, and functions $M_{ij}(x,t)$. In this section we shall prove some comparison theorems for such a problem B and a second problem B^* involving corresponding quantities $\omega^*(x,\eta,\eta')$, $\Phi_\alpha(x,\eta,\eta')$ $(\alpha = 1, \cdots, m)$, $Q^*[\eta(a),\eta(b)]$, $\Psi^*_\nu[\eta(a),\eta(b)]$ $(\nu = 1, \cdots, p^*)$, $M^*_{ij}(x,t)$. The expressions (1.1) for B and B^* will be denoted by $I[\eta]$ and $I^*[\eta]$, respectively. We shall assume that B and B^* involve the same auxiliary differential equations $\Phi_\alpha = 0$ $(\alpha = 1, \cdots, m)$, and that each of these problems satisfies hypotheses (H) of Section 2. The set of characteristic values and corresponding characteristic solutions of B and B^* will be denoted by $\lambda_s, \eta_i = y_{is}(x)$, $\zeta_i = z_{is}(x)$ and λ^*_s, $\eta_i = y^*_{is}(x)$, $\zeta_i = z^*_{is}(x)$ $(s = 1, 2, \cdots)$, respectively. Each of the sets of characteristic values is supposed to be ordered as in (4.7), and for each of the problems the characteristic solutions are chosen orthonormal in the sense of (4.8). The classes of admissible arcs $\mathfrak{S}_s, \mathfrak{S}^*_s$ $(s = 1, 2, \cdots)$ are defined for the two problems as described in the preceding section.

As indicated in the introduction, when $M_{ij}(x,t) \equiv 0$, the theorems of this section are essentially those previously given by Morse [16] and Hu [9]. So far as the author knows, however, for even the simplest type of integro-differential system of the form here considered when the functions M_{ij} are not all identically zero, these theorems are new. As indicated by the examples at the end of Section 4, integro-differential systems of the type here considered do not in general possess some of the well known properties of differential systems. Hence the fact that such integro-differential systems do possess the same type of comparison and oscillation theorems as corresponding differential boundary problems is of significance. In Section 7 the results of the present section will be used to prove corresponding theorems for integro-differential systems involving the characteristic parameter in a more general fashion. It is to be emphasized that in the proof of the results of this and the following section, *no assumption of normality on sub-intervals is made.*

We shall first prove the following general comparison theorem.

THEOREM 5.1. *Suppose that B and B^* have in common $\Phi_\alpha(x,\eta,\eta')$ $(\alpha=1,\cdots,m)$, $\Psi_\gamma[\eta(a),\eta(b)]$ $(\gamma=1,\cdots,p)$, and that $I^*[\eta]\geqq I[\eta]$ for arbitrary admissible arcs η. Then $\lambda^*_s\geqq\lambda_s$ $(s=1,\cdots)$.*

Corresponding to a positive integer s, let $d_1,\cdots,d_s$ be real constants such that $d_1^2+\cdots+d_s^2=1$, and the arc $y_i=y^*_{i1}d_1+\cdots+y^*_{is}d_s$ satisfies the relations $K[y_\kappa;\eta]=0$ $(\kappa=1,\cdots,s-1)$. The orthonormal relations satisfied by the characteristic solutions of B^* imply $K[\eta]=1$, and hence η belongs to $\mathfrak{S}_s$. The inequalities

$$\lambda^*_s\geqq I^*[\eta]\geqq I[\eta]\geqq\lambda_s$$

are then consequences of Theorem 4.3 and the minimizing property of λ_s.

The following corollary is immediate

COROLLARY 1. *If the hypotheses of Theorem 5.1 are strengthened so that $I^*[\eta]>I[\eta]$ for arbitrary non-identically vanishing admissible arcs η, then $\lambda^*_s>\lambda_s$ $(s=1,2,\cdots)$.*

COROLLARY 2. *For a given problem B, let B^0 denote a particular related problem involving the same end-form Q and end conditions $\Psi_\gamma=0$ $(\gamma=1,\cdots,p)$, for which $M^0_{ij}(x,t)\equiv 0$, and $2\omega^0[x,\eta,\eta']\equiv 2\omega(x,\eta,\eta')-l_0\eta_i\eta_i$, where l_0 is a constant such that $M[\eta]\leqq l_0K[\eta]$ for arbitrary continuous arcs $\eta\equiv[\eta_i(x)]$. If $\{\lambda_s^0\}$ denote the ordered set of characteristic values of B^0, then $\lambda_s\geqq\lambda_s^0$ $(s=1,2,\cdots)$.*

Suppose two problems B and B^* have in common $\Phi_\alpha(x,\eta,\eta')$ $(\alpha=1,\cdots,m)$ and $\Psi_\gamma[\eta(a),\eta(b)]$ $(\gamma=1,\cdots,p)$. We shall denote by the difference problem[6] $D(B,B^*)$ the problem involving these same relations, and the expression $\Delta I[\eta]=I^*[\eta]-I[\eta]$. Hypotheses (H_1) and (H_3) for $D(B,B^*)$ are consequences of these hypotheses for B and B^*.

THEOREM 5.2. *Suppose hypotheses (H) are satisfied by each of the problems B, B^*, $D(B,B^*)$. Let $\lambda=\boldsymbol{\lambda}_s$, $\eta_i=\mathbf{y}_{is}(x)$, $\zeta_i=\mathbf{z}_{is}(x)$ $(s=1,2,\cdots)$ denote the characteristic values and characteristic solutions of $D(B,B^*)$, supposedly ordered as in (4.7) and orthonormal in the sense of (4.8). If s and t are arbitrary positive integers, then $\lambda^*_{s+t-1}\geqq\lambda_s+\boldsymbol{\lambda}_t$.*

---

[6] This terminology is due to Morse [16], p. 100.

Let $d_1, \cdots, d_{s+t-1}$ be real constants such that the arc

$$\eta_i = y^*_{i1} d_1 + \cdots + y^*_{is+t-1} d_{s+t-1}$$

satisfies the $s + t - 2$ conditions

$$K[y_\kappa; \eta] = 0 \quad (\kappa = 1, \cdots, s-1), \qquad K[\mathbf{y}_\nu; \eta] = 0 \quad (\nu = 1, \cdots, t-1),$$

and the relation $K[\eta] = 1$. Then by Theorem 4.3 and the minimizing properties of λ_s and $\boldsymbol{\lambda}_t$ we have

$$\lambda^*_{s+t-1} \geqq I^*[\eta] = I[\eta] + \Delta I[\eta] \geqq \lambda_s + \boldsymbol{\lambda}_t.$$

COROLLARY. *Under the hypotheses of Theorem 5.2, if h_s denote the number of characteristic values of $D(B, B^*)$ less than λ_s, then $\lambda^*_{s+h_s} \geqq 2\lambda_s$.*

We shall now consider two problems B and B^* which differ only in the end-forms Q and Q^*. Let

$$\begin{aligned} 2Q^* - 2Q &\equiv 2d[\eta(a), \eta(b)] \\ &= d_{ia;ja}\eta_i(a)\eta_j(a) + 2d_{ia;jb}\eta_i(a)\eta_j(b) + d_{ib;jb}\eta_i(b)\eta_j(b). \end{aligned}$$

The problems B and B^* involve the same system of integro-differential equations. Let r denote the order of anormality on ab of this system, as defined in Section 2, and suppose $\eta_i \equiv 0$, $\zeta_i = v_{ih}(x)$ $(h = 1, \cdots, r)$ are r linearly independent solutions of this system on ab. The comparison theorem to be established involves the number of positive and negative zeros of the determinant

$$D(\rho) = \begin{vmatrix} d_{ia;jb} - \rho\delta_{ij} & d_{ia;jb} & \Psi_{\nu;ia} & -v_{ik}(a) \\ d_{ja;ib} & d_{ib;jb} - \rho\delta_{ij} & \Psi_{\nu;ib} & v_{ik}(b) \\ \Psi_{\gamma;ja} & \Psi_{\gamma;jb} & O_{\gamma\nu} & O_{\gamma k} \\ -v_{jh}(a) & v_{jh}(b) & O_{h\nu} & O_{hk} \end{vmatrix}, \tag{5.1}$$

where $i, j = 1, \cdots, n;\ \gamma, \nu = 1, \cdots, p;\ h, k = 1, \cdots, r$.

The determinant $D(\rho)$ is a polynomial in ρ of degree $2n - p - r$. It is well known (see, for example, Caratheodory [IV], pp. 164-189) that the zeros of $D(\rho)$ are all real, and that if a given value is a zero of $D(\rho)$ of multiplicity q there are exactly q linearly independent solutions of the linear homogeneous equations in $2n + p + r$ unknowns whose coefficients are the elements of the rows of the matrix of $D(\rho)$. Let $\rho_1 \leqq \rho_2 \leqq \cdots \leqq \rho_{2n-p-r}$ denote these zeros of $D(\rho)$, and let $w = w_{lg}$ $(l = 1, \cdots, 2n + p + r;\ g = 1, \cdots, 2n - p - r)$ be a solution of this system of algebraic equations corresponding to $\rho = \rho_g$. The solutions may be chosen orthonormal in the sense that $w_{\sigma g} w_{\sigma g'} = \delta_{gg'}$

$(g, g' = 1, \cdots, 2n - p - r;\ \sigma = 1, \cdots, 2n)$. Moreover, if R is a given value such that $0 \leq R < 2n - p - r$, and η_{ia}, η_{ib} is a set of values such that

$$(5.2) \qquad \begin{array}{ll} \Psi_\gamma[\eta_a; \eta_b] = 0,\ -v_{ih}(a)\eta_{ia} + v_{ih}(b)\eta_{ib} = 0 & \\ & (\gamma = 1, \cdots, p;\ h = 1, \cdots, r), \\ w_{ig}\eta_{ia} + w_{n+i\, g}\,\eta_{ib} = 0 & (g = 1, \cdots, R), \end{array}$$

then $2d[\eta_a; \eta_b] \geq \rho_{R+1}[\eta_{ia}\eta_{ia} + \eta_{ib}\eta_{ib}]$.

We now prove the following comparison theorem:

THEOREM 5. 3. *Suppose the boundary problems B and B^* differ only in the end-forms Q and Q^*. Let N and P denote, respectively, the number of negative and positive zeros of $D(\rho)$ defined by* (5. 1) *in terms of the difference end-form $d = Q^* - Q$, each zero being counted a number of times equal to its multiplicity. Then for an arbitrary positive integer s, $\lambda^*_{s+N} \geq \lambda_s$ and $\lambda_{s+P} \geq \lambda^*_s$.*

There clearly exist constants $d_1, \cdots, d_{s+N}$ such that $d_1^2 + \cdots + d^2_{s+N} = 1$, the arc $\eta_i = y^*_{i1}d_1 + \cdots + y^*_{is+N}d_{s+N}$ satisfies the relations $K[y_\kappa; \eta] = 0$ $(\kappa = 1, \cdots, s - 1)$, and the set of end-values $\eta_i(a)$, $\eta_i(b)$ is orthogonal to each of the sets $[w_{\sigma g}]$ $(g = 1, \cdots, N)$ in the sense of (5. 2). Since the arc η is admissible, its end-values also satisfy the other relations of (5. 2), and consequently

$$(5.3) \qquad 2d[\eta(a), \eta(b)]) \geq \rho_{N+1}[\eta_i(a)\eta_i(a) + \eta_i(b)\eta_i(b)] \geq 0.$$

Finally, since $K[\eta] = 1$, the arc η belongs to $\mathfrak{S}_s$ and by Theorem 4. 3 and the minimum property of λ_s we have

$$\lambda^*_{s+N} \geq I^*[\eta] = I[\eta] + 2d[\eta(a), \eta(b)] \geq I[\eta] \geq \lambda_s.$$

The remainder of the conclusion of the theorem is a ready consequence of the relation $I[\eta] = I^*[\eta] - 2d[\eta(a), \eta(b)]$, together with the fact that when the difference form d is replaced by its negative the zeros of $D(\rho)$ are also replaced by their negatives.

If Δ is a finite interval, either closed or open, of the λ-axis, we shall denote by $V[\Delta]$ the number of characteristic values of B on Δ. Corresponding to a value L we shall also denote by $V_L(W_L)$ the number of characteristic values of B which are less (not greater) than L. The numbers $V^*[\Delta]$, V^*_L, W^*_L are defined for B^* in an analogous manner.

The following corollary is an immediate consequence of the above theorem.

COROLLARY. *Under the hypotheses of Theorem* 5. 3, *for every L we have*

$V_L - P \leq V^*_L \leq V_L + N$, $W_L - P \leq W^*_L \leq W_L + N$; moreover, $|V[\Delta] - V^*[\Delta]| \leq N + P$ for every finite sub-interval Δ of the λ-axis.

Now suppose that two problems B and B^* satisfy hypotheses (H), and differ only in the end-conditions $\Psi_\gamma[\eta(a), \eta(b)] = 0$ $(\gamma = 1, \cdots, p)$ and $\Psi^*_\nu[\eta(a), \eta(b)] = 0$ $(\nu = 1, \cdots, p^*)$. Corresponding to the terminology of Morse ([16], p. 92), we shall say that B^* is a sub-problem of B if the matrix

$$(5.4) \qquad \left\| \begin{matrix} \Psi_{\gamma;\, ia} & \Psi_{\gamma;\, jb} \\ \Psi^*_{\nu;\, ja} & \Psi^*_{\nu;\, jb} \\ -v_{jh}(a) & v_{jh}(b) \end{matrix} \right\| (\gamma = 1, \cdots, p;\ \nu = 1, \cdots, p^*;\ h = 1, \cdots, r)$$

has rank $r + p^*$. If B^* is a sub-problem of B, clearly $p^* \geq p$. The number $p^* - p$ will be called the dimension of B^* as a sub-problem of B. It follows readily that if B^* is a sub-problem of B and η is an admissible arc for B^*, then η is also an admissible arc for B. Moreover, since (5.4) has rank $r + p^*$, if B^* is a sub-problem of B then in the class of differentially admissible arcs η the end-conditions $\Psi^*_\nu = 0$ $(\nu = 1, \cdots, p^*)$ are equivalent to the p conditions $\Psi_\gamma = 0$ $(\gamma = 1, \cdots, p)$ together with $p^* - p$ other end-conditions.

In view of the above remarks the following results are consequences of the minimizing properties of the characteristic values of B and B^*, together with the result of Theorem 4.3. Since the details of proof are the same as those used by Hu ([19], Section 15) to prove the corresponding results for the differential system he considered, these results will be stated without proof.

THEOREM 5.4. *If B^* is a sub-problem of B of dimension $p^* - p$, then $\lambda_{s+p^*-p} \geq \lambda^*_s \geq \lambda_s$ $(s = 1, 2, \cdots)$.*

COROLLARY. *Under the hypotheses of Theorem 5.4, for every L we have $V_L - (p^* - p) \leq V^*_L \leq V_L$, $W_L - (p^* - p) \leq W^*_L \leq W_L$; moreover, $|V[\Delta] - V^*[\Delta]| \leq p^* - p$ for every finite sub-interval Δ of the λ-axis.*

Let us now consider two problems B and B^* which have in common $\omega(x, \eta, \eta')$, $\Phi_\alpha(x, \eta, \eta')$, $M_{ij}(x, t)$ but involving, respectively, the sets

$$Q[\eta(a), \eta(b)], \qquad \Psi_\gamma[\eta(a), \eta(b)] \qquad (\gamma = 1, \cdots, p)$$

and

$$Q^*[\eta(a), \eta(b)], \qquad \Psi^*_\nu[\eta(a), \eta(b)] \qquad (\nu = 1, \cdots, p^*).$$

We suppose, as before, that each of the problems B and B^* satisfies hypotheses (H). Let $q + r$ denote the rank of the matrix (5.4). Clearly, $q \geq p$,

$q \geq p^*$, $q \leq p + p^*$. There are seen to exist q end-relations $\Psi_\mu^0[\eta(a), \eta(b)]$ $(\mu = 1, \cdots, q)$ satisfying (H_3) and such that the problem B_0 involving ω, Φ_a, M_{ij}, Q, Ψ_μ^0 is a sub-problem of B, and the problem B_0^* involving ω, Φ_a, M_{ij}, Q^*, Ψ_μ^0 is a sub-problem of B^*. For brevity we shall say that the set of end-conditions $\Psi_\mu^0 = 0$ $(\mu = 1, \cdots, q)$ is the intersection of the sets $\Psi_\gamma = 0$ $(\gamma = 1, \cdots, p)$ and $\Psi^*_\nu = 0$ $(\nu = 1, \cdots, p^*)$ relative to the differential equations $\Phi_a = 0$. In comparing the two problems B_0 and B^*_0 use is made of the determinant (5.1) for these two problems. This determinant will be denoted by $D_0(\rho)$, and is seen to be a polynomial of degree $2n - r - q$ in ρ.

The following comparison theorem is the analogue of a theorem proved by Morse ([16], p. 94) for the differential system he considered, and is a consequence of the preceding theorems of this section.

THEOREM 5.5. *Suppose B and B^* have in common ω, Φ_α $(\alpha = 1, \cdots, m)$, $M_{ij}(x, t)$ and involve, respectively, the sets Q, Ψ_γ $(\gamma = 1, \cdots, p)$ and Q^*, Ψ^*_ν $(\nu = 1, \cdots, p^*)$. Suppose, moreover, that $\Psi_\mu^0 = 0$ $(\mu = 1, \cdots, q)$ is the intersection of $\Psi_\gamma = 0$ $(\gamma = 1, \cdots, p)$ and $\Psi^*_\nu = 0$ $(\nu = 1, \cdots, p^*)$ relative to the differential equations $\Phi_\alpha = 0$, and denote by N_0 and P_0, respectively, the number of negative and positive zeros of the determinant $D_0(\rho)$ defined above. Then $\lambda^*_{s+N_0+q-p^*} \geq \lambda_s$, $\lambda_{s+P_0+q-p} \geq \lambda^*_s$ $(s = 1, 2, \cdots)$.*

COROLLARY. *Under the hypotheses of Theorem 5.5, for every L we have*

$$V_L - P_0 - (q - p) \leq V^*_L \leq V_L + N_0 + (q - p^*),$$
$$W_L - P_0 - (q - p) \leq W^*_L \leq W_L + N_0 + (q - p^*);$$

moreover, $| V[\Delta] - V^*[\Delta] | \leq N_0 + P_0 + 2q - p - p^*$ for every finite subinterval Δ of the λ-axis, and $N_0 + P_0 + 2q - p - p^* \leq 2n - r$.

The last conclusion of this corollary is a ready consequence of the inequalities $N_0 + P_0 \leq 2n - q - r$, $q \leq p + p^*$.

6. Oscillation theorems. Suppose that $\omega(x, \eta, \eta')$, $\Phi_\alpha(x, \eta, \eta')$ $(\alpha = 1, \cdots, m)$, $M_{ij}(x, t)$ are given and satisfy on $a \leq x \leq b$ hypotheses (H_1) and (H_2); moreover, that the end-form Q depends only upon $\eta_i(a)$, that is, $2Q \equiv Q_{ia;ja}\eta_i(a)\eta_j(a)$. Let $\| \mathfrak{P}_{i\rho} \|$ $(\rho = 1, \cdots, q \leq n;\ i = 1, \cdots, n)$ be a constant matrix of rank q, and denote by P the linear sub-space

$$P: x = a, \quad \eta_{ia}\mathfrak{P}_{i\rho} = 0 \qquad (\rho = 1, \cdots, q) \tag{6.1}$$

in $(n + 1)$-dimensional (x, η_{ia}) space. For a given c on $a < c \leq b$, let $B(P:c)$ denote the normal boundary problem determined by ω, Φ_α, $M_{ij}(x, t)$, $Q[\eta(a)]$; and the end conditions

$$(6.2) \qquad \eta_i(a)\mathfrak{P}_{i\rho} = 0,\ \eta_i(c) = 0 \qquad (\rho = 1, \cdots, q;\ i = 1, \cdots, n).$$

Corresponding to a value c on $a < c \leq b$, denote by r_c the order of anormality on ac, and by $\eta_i = \eta_{ih}(x) \equiv 0$, $\zeta_i = \zeta_{ih}(x)$ $(h = 1, \cdots, r_c)$ a linearly independent set of solutions of the integro-differential equations of $B(P:c)$ on ac. If the matrix $\| \mathfrak{P}_{i\rho}; \zeta_{ih}(a) \|$ $(\rho = 1, \cdots, q; h = 1, \cdots, r_c; i = 1, \cdots, n)$ has rank $q + r_c - k_c$, then $0 \leq k_c \leq q$ and there are values $\rho_1, \cdots, \rho_{l(c)}$ $(l(c) = q - k_c)$ such that $1 = \rho_1 < \rho_2 < \cdots < \rho_{l(c)} \leq q$ and the auxiliary end conditions $\Psi_\gamma = 0$ $(\gamma = 1, \cdots, p(c) = n + q - k_c)$ of $B(P:c)$ may be written

$$(6.3) \quad \eta_i(a)\mathfrak{P}_{iv} = 0,\ \eta_i(c) = 0 \qquad (v = \rho_1, \rho_2, \cdots, \rho_{l(c)};\ i = 1, \cdots, n).$$

It is to be remarked that the coefficients of the end conditions of $B(P:c)$ may be chosen the same for all values of c on an interval where r_c is constant.

The expression (1.1) for $B(P:c)$ will be denoted by $I[\eta:c]$. It is to be emphasized that the upper limit of the integrals in $I[\eta:c]$ is c. If $a < c < \mathbf{c} \leq b$, and $\eta_i = u_i(x)$ is an admissible arc for $B(P:c)$, then the arc $U_i = u_i(x)$ $(a \leq x \leq c)$, $U_i \equiv 0$ $(c \leq x \leq \mathbf{c})$ is admissible for $B(P:\mathbf{c})$ and $I[U:\mathbf{c}] = I[U:c] = I[u:c]$.

A value c will be called a *focal point of P on the interval ab relative to the system*

$$(6.4) \qquad L_i[\eta, \zeta] = 0, \quad L_{n-i}[\eta, \zeta|\lambda] = 0$$

for $\lambda = \lambda_0$ *if:* (i) $\lambda = \lambda_0$ *is a characteristic value of* $B(P:c)$; (ii) *there is at least one corresponding characteristic solution* $\eta_i = y_i(x)$, $\zeta_i = z_i(x)$ *such that for arbitrary* ϵ *satisfying* $0 < \epsilon < b - c$ *there is no corresponding set of functions* $\zeta_i(x)$ *forming with the arc* $\eta_i = y_i(x)$ *on ac,* $\eta_i \equiv 0$ *on* $c \leq x \leq c + \epsilon$ *a characteristic solution of* $B(P:c+\epsilon)$ *for* $\lambda = \lambda_0$. If $a < c < b$ and c is a focal point of P on ab relative to (6.4) for $\lambda = \lambda_0$, then c is a focal point of P on every interval ab' $(c < b' \leq b)$ relative to the same system. On the other hand, for $c = b$ condition (ii) of the above definition is satisfied vacuously, and b is a focal point of P on ab relative to the system (6.4) for $\lambda = \lambda_0$ whenever λ_0 is a characteristic value of $B(P:b)$. Consequently, in general $x = b$ may be a focal point of P on ab relative to (6.4) for $\lambda = \lambda_0$ and not be a focal point of P on ab'' $(b'' > b)$ relative to this same system. If the integro-differential system (6.4) is normal on every sub-interval of an extension ab'' $(b'' > b)$, then a point c of $a < c \leq b$ is a focal point on the interval ab'' relative to (6.4) for $\lambda = \lambda_0$ if and only if λ_0 is a characteristic value of $B(P:c)$. If y_i, z_i is a solution of $B(P:c)$ satisfying the above condition (ii), we shall say that $y \equiv [y_i(x)]$ is an arc *determining* $x = c$ *as* *a focal point of P.* The number of such solutions of $B(P:c)$ which are linearly independent on ac will be termed the *index of c as a focal point.* If the linear space (6.1) reduces to the point $x=a$, $\eta_{ia}=0$, that is, if $q=n$, the corresponding focal points will be termed *conjugate points* of $x=a$.

We shall denote by $\{\lambda_s(c)\}$, $\eta_i=y_{is}(x:c)$, $\zeta_i=z_{is}(x:c)$ $(s=1,2,\cdots)$ the characteristic values and characteristic solutions of $B(P:c)$, supposed ordered and orthonormal in the sense of (4.7) and (4.8).

LEMMA 6.1. *As* $c\rightarrow a$, $\lambda_1(c)\rightarrow+\infty$.

In view of Corollary 2 of Theorem 5.1 it is sufficient to prove this theorem for the associated differential problem $B^0(P:c)$. For $B^0(P:c)$ the desired result may be established by the method used by Hu ([9], Lemma 13.1), since his proof does not involve any assumption of normality.

LEMMA 6.2. *Each of the characteristic values* $\lambda_s(c)$ *of* $B(P:c)$ *varies continuously with c, and increases from* $\lambda_s(b)$ *to* $+\infty$ *as c decreases from b to a.*

Suppose $a<c\leq\boldsymbol{c}\leq b$, and let $d_1,\cdots,d_s$ be constants such that the arc $\eta_i=y_{i1}(x:c)d_1+\cdots+y_{is}(x:c)d_s$ on ac, $\eta_i\equiv0$ on $c\boldsymbol{c}$ satisfies the relations $1=K[\eta:c]=K[\eta:\boldsymbol{c}]$; $K[y_\kappa(x:\boldsymbol{c}),\eta:\boldsymbol{c}]=0$, $(\kappa=1,\cdots,s-1)$. The arc η thus defined is admissible for both $B(P:c)$ and $B(P:\boldsymbol{c})$. By Theorem 4.3 and the minimizing property of $\lambda_s(\boldsymbol{c})$ we have

$$\lambda_s(\boldsymbol{c})\leq I[\eta:\boldsymbol{c}]=I[\eta:c]\leq\lambda_s(c),$$

that is, $\lambda_s(c)$ is a monotone decreasing function on $a<c\leq b$.

We shall now prove that $\lambda_s(c)$ is continuous on $a<c\leq b$. Suppose $a<\boldsymbol{c}<b$, and denote by $\boldsymbol{\Psi}_\gamma[\eta(a),\eta(\boldsymbol{c})]=0$ $(\gamma=1,\cdots,p(\boldsymbol{c}))$ the corresponding auxiliary boundary conditions (6.3) for $B(P:\boldsymbol{c})$. Let $\epsilon_1>0$ be such that for $a<\boldsymbol{c}-\epsilon_1\leq c\leq\boldsymbol{c}+\epsilon_1<b$ hypothesis (H_3) is satisfied by the end conditions $\boldsymbol{\Psi}_\gamma[\eta(a),\eta(c)]=0$ $(\gamma=1,\cdots,p(\boldsymbol{c}))$, and, moreover, ϵ_1 is such that r_c is constant on $\boldsymbol{c}-\epsilon_1\leq c\leq\boldsymbol{c}$. For each c on this interval let $\boldsymbol{B}(P:c)$ denote the boundary problem determined by $I[\eta:c]$, Φ_a, $\boldsymbol{\Psi}_\gamma$. The ordered characteristic values and orthonormal characteristic solutions of $\boldsymbol{B}(P:c)$ will be denoted by $\boldsymbol{\lambda}_s(c)$, $\boldsymbol{y}_{is}(x:c)$, $\boldsymbol{z}_{is}(x:c)$ $(s=1,2,\cdots)$. Since the order of anormality on ac is constant for $\boldsymbol{c}-\epsilon_1\leq c\leq\boldsymbol{c}$, for such values of c the problem $\boldsymbol{B}(P:c)$ is identical with $B(P:c)$. If r_c is also constant on $\boldsymbol{c}\leq c\leq\boldsymbol{c}+\epsilon_1$, $\boldsymbol{B}(P:c)$ is also identical with $B(P:c)$ on this interval. If, however, r_c has a discontinuity at $\boldsymbol{c}$, then $B(P:c)$ is a sub-problem of $\boldsymbol{B}(P:c)$ on $\boldsymbol{c}\leq c\leq\boldsymbol{c}+\epsilon_1$ and, by Theorem 5.4, $\lambda_s(c)\geq\boldsymbol{\lambda}_s(c)$ $(s=1,2,\cdots)$.

It will now be proved that each $\boldsymbol{\lambda}_s(c)$ is continuous at $c = \boldsymbol{c}$. Let $\boldsymbol{s}_\sigma[\eta, \zeta : c] = 0$ $(\sigma = 1, \cdots, 2n)$ denote the boundary conditions of $\boldsymbol{B}(P : c)$. By the Corollary to Lemma 3. 2 there is a λ_0 such that for c on $| c - \boldsymbol{c} | \leqq \epsilon_1$ the differential boundary problem

(6. 5) $\boldsymbol{L}_i[\eta,\zeta] \equiv L_i[\eta,\zeta] = 0,\ \boldsymbol{L}_{n+i}[\eta,\zeta] \equiv \zeta'_i - \Omega_{\eta_i} + \lambda_0\eta_i = 0,\ \boldsymbol{s}_\sigma[\eta,\zeta : c] = 0$

is incompatible. Suppose $\boldsymbol{c} - \epsilon_1 \leq c \leq c' \leq \boldsymbol{c} + \epsilon_1$, and consider the nonhomogeneous system

(6. 6) $\boldsymbol{L}_i[\eta, \zeta] = 0,\ \boldsymbol{L}_{n+i}[\eta, \zeta] = h_{if}(x : c'),\ \boldsymbol{s}_\sigma[\eta, \zeta : c] = 0 \ \ (f = 1, \cdots, s),$

where

$$h_{if}(x : c') = [\lambda_0 - \boldsymbol{\lambda}_f(c')]\boldsymbol{y}_{if}(x : c') + \int_a^{c'} M_{ij}(x, t)\boldsymbol{y}_{if}(t : c')\,dt.$$

If $\eta_{if}(x : c, c')$, $\zeta_{if}(x : c, c')$ is the solution of (6. 6), we have

$$\eta_{if}(x : c', c') = \boldsymbol{y}_{if}(x : c'), \qquad \zeta_{if}(x : c', c') = \boldsymbol{z}_{if}(x : c').$$

For arbitrary constants d_f $(f = 1, \cdots, s)$, set

$$I[d : c, c'] = I[\eta_f(x : c, c')d_f : c], \qquad K[d : c, c'] = K[\eta_f(x : c, c')d_f : c].$$

Using the explicit form of $\eta_{if}(x : c, c')$, $\zeta_{if}(x : c, c')$ as given by the Green's matrix of (6. 5), together with Lemma 3. 3, it follows that the coefficients of the quadratic forms $I[d : c, c']$, $K[d : c, c']$ in the variables d_f are continuous functions of c, c' uniformly on $\boldsymbol{c} - \epsilon_1 \leq c \leq c' \leq \boldsymbol{c} + \epsilon_1$. Since

$$I[d : c', c'] = \sum_{f=1}^{s} \boldsymbol{\lambda}_f(c')d_f^2, \qquad K[d : c', c'] = \sum_{f=1}^{s} d_f^2,$$

for a given ϵ $(0 < \epsilon < 1)$ there is a $\delta_\epsilon < \epsilon$ such that if $\boldsymbol{c} - \epsilon_1 \leq c \leq c' \leq \boldsymbol{c} + \epsilon_1$, $c' - c \leq \delta_\epsilon$, then

$$K[d : c, c'] \geq (1 - \epsilon)d_f d_f,\ I[d : c, c'] \leq [\boldsymbol{\lambda}_s(c') + \epsilon] \cdot K[d : c, c'].$$

This implies, in particular, that the functions $\eta_{if}(x : c, c')$ are linearly independent on ac. Now the constants d_f may be chosen so that the arc $\eta_i = \eta_{if}(x : c, c')d_f$ satisfies the relations $K[y_\kappa(x : c) ; \eta : c] = 0$, $(\kappa = 1, \cdots, s - 1)$, $K[\eta : c] = 1$, and for such a choice we have $\boldsymbol{\lambda}_s(c) \leq I[\eta : c] \leq \boldsymbol{\lambda}_s(c') + \epsilon$. In particular, if $0 \leq \delta \leq \delta_\epsilon$ we have

$$\lambda_s(\boldsymbol{c}) + \epsilon = \boldsymbol{\lambda}_s(\boldsymbol{c}) + \epsilon \geq \boldsymbol{\lambda}_s(\boldsymbol{c} - \delta) = \lambda_s(\boldsymbol{c} - \delta) \geq \lambda_s(\boldsymbol{c}),$$
$$\lambda_s(\boldsymbol{c}) - \epsilon = \boldsymbol{\lambda}_s(\boldsymbol{c}) - \epsilon \leq \boldsymbol{\lambda}_s(\boldsymbol{c} + \delta) \leqq \lambda_s(\boldsymbol{c} + \delta) \leq \lambda_s(\boldsymbol{c}),$$

that is, $| \lambda_s(\boldsymbol{c} \pm \delta) - \lambda_s(\boldsymbol{c}) | \leq \epsilon$. Since ϵ may be chosen arbitrarily small we have proved that $\lambda_s(c)$ is continuous at $c = \boldsymbol{c}$, whenever $a < \boldsymbol{c} < b$. The continuity of $\lambda_s(c)$ at $c = b$ is proved by the same method, where in the above formulas $\boldsymbol{c} = b$, and the comparison values c are restricted by the inequality $b - \epsilon_1 \leq c \leq b$.

THEOREM 6.1. *The number of focal points of P on $a < x < b$ relative to the system* (6.4) *for a given value $\lambda = \lambda^*$ is equal to the number of characteristic values $\lambda_s(b)$ of $B(P:b)$ which are less than λ^*.*

Suppose that there are exactly s characteristic values $\lambda_f(b)$ $(f = 1, \cdots, s)$ of $B(P:b)$ which are less than λ^*, and for each value of f let $x = c_f$ denote the last value on $a < x < b$ for which $\lambda_f(x) = \lambda^*$. In view of Lemmas 6.1 and 6.2 these values c_f are well-defined. It will first be proved that if $a < c < b$ and $x = c$ is a focal point of P relative to (6.4) for $\lambda = \lambda^*$, then c is one of the values c_f. For suppose c is such a focal point which is distinct from the values c_f, and denote by $y \equiv [y_i(x)]$ an arc determining c as a focal point of P. We may without loss of generality assume $K[y:c] = 1$. Let N denote the largest integer such that $\lambda_N(c) < \lambda^*$. Then $\lambda_{N+1}(c) = \lambda^*$, and since c is distinct from the values c_f there exists an $\epsilon > 0$ such that $\lambda_{N+1}(x) \equiv \lambda^*$ on $c \leq x \leq c + \epsilon$. Now consider an arc η of the form

$$\eta_i = y_{i1}(x:c)d_1 + \cdots + y_{iN}(x:c)d_N + y_i(x)d_{N+1} \quad (a \leq x \leq c),$$

and $\eta_i \equiv 0$ on $c \leq x \leq c + \epsilon$, where the constants $d_1, \cdots, d_{N+1}$ are such that

$$(6.7) \quad 1 = K[\eta : c + \epsilon] = K[\eta : c], \quad K[y_\kappa(x : c + \epsilon) ; \eta : c + \epsilon] = 0 \quad (\kappa = 1, \cdots, N).$$

In view of the minimizing property of $\lambda_{N+1}(c + \epsilon)$ we have $I[\eta : c + \epsilon] \geq \lambda_{N+1}(c + \epsilon)$. By Theorem 4.3, however,

$$I[\eta : c + \epsilon] = I[\eta : c] \leq \lambda_{N+1}(c) = \lambda_{N+1}(c + \epsilon).$$

Hence $I[\eta : c + \epsilon] = \lambda_{N+1}(c + \epsilon)$, $0 = d_1 = \cdots = d_N$ by (4.9), and in view of (6.7) and Theorem 4.4 there exist functions ζ_i such that η_i, ζ_i is a characteristic solution of $B(P : c + \epsilon)$. This, however, contradicts the assumption that y is an arc determining c as a focal point of P. We have established, therefore, that if c is a focal point of P on $a < x < b$ relative to (6.4) for $\lambda = \lambda^*$, then c is one of the values c_f $(f = 1, \cdots, s)$ defined above.

It will now be proved that each c_f $(f = 1, \cdots, s)$ is a focal point of P relative to (6.4) for $\lambda = \lambda^*$, and that the index of c_f is equal to the number of characteristic values $\lambda_\kappa(c_f)$ for which $\kappa \leq s$ and $\lambda_\kappa(c_f) = \lambda^*$. It is obviously sufficient to prove this result for the largest value c_s, since each c_f may be considered the largest such value on a suitable subinterval of ab. Suppose that $\lambda^* = \lambda_s(c_s) = \lambda_{s-1}(c_s) = \cdots = \lambda_{s-g+1}(c_s)$, and either $s - g = 0$ or $\lambda_{s-g}(c_s) < \lambda^*$. It is also quite possible that some of the values $\lambda_\kappa(c_s)$ with $\kappa > s$ are also equal to λ^*; if this is the case, however, the corresponding $\lambda_\kappa(c)$ is constant on $c_s \leq c \leq b$. We shall suppose, for definiteness, that there do exist values $\lambda_\kappa(c_s) = \lambda^*$ with $\kappa > s$, and shall give the proof of the theorem in this more complicated case. If $\lambda_{s+1}(c_s) > \lambda^*$ the theorem is proved in a similar manner, but with obvious simplifications. If

$$\lambda^* = \lambda_{s+1}(c_s) = \cdots = \lambda_{s+t}(c_s), \qquad \lambda_{s+t+1}(c_s) > \lambda^*,$$

set

$$\eta_i = y_{i1}(x: c_s)d_1 + \cdots + y_{i.s+t}(x: c_s)d_{s+t}$$

on ac_s, $\eta_i \equiv 0$ on $c_s b$, where the constants $d_1, \cdots, d_{s+t}$ are chosen so that

$$(6.8) \quad K[y_f(x: b); \eta: b] = 0 \quad (f = 1, \cdots, s),$$
$$1 = d_1^2 + \cdots + d^2_{s+t} = K[\eta: c_s] = K[\eta: b].$$

From the minimizing property of $\lambda_{s+1}(b)$ we have $I[\eta: b] \geq \lambda_{s+1}(b) = \lambda^*$. On the other hand, by Theorem 4.3, $I[\eta: b] = I[\eta: c_s] \leq \lambda_{s+t}(c_s) = \lambda^*$. By Theorem 4.4 there are functions ζ_i such that η_i, ζ_i is a characteristic solution of $B(P: b)$ for $\lambda = \lambda^*$. Since the functions $y_{i\kappa}(x: c_s)$ $(\kappa = 1, \cdots, s + t)$ are linearly independent on ac_s, there are seen to exist at least t linearly independent solutions η_{il}, ζ_{il} $(l = s + 1, \cdots, s + t)$ of $B(P: b)$ for $\lambda = \lambda^*$ with $\eta_{il} \equiv 0$ on $c_s b$. Now η_{il}, ζ_{il} $(l = s + 1, \cdots, s + t)$ are linearly independent solutions of $B(P: c_s)$ for $\lambda = \lambda^*$ such that no arc $\eta_{il}(x)$ is an arc determining c_s as a focal point of P. Consequently, the index of c_s as a focal point is at most g. On the other hand, there are g solutions $\eta_{i\kappa}, \zeta_{i\kappa}$ $(\kappa = s - g + 1, \cdots, s)$ of $B(P: c_s)$ for $\lambda = \lambda^*$ satisfying the relations $K[\eta_\kappa; \eta_\mu: c_s] = \delta_{\kappa\mu}$, $K[\eta_l, \eta_\kappa: c_s] = 0$ $(\kappa, \mu = s - g + 1, \cdots, s;$ $l = s + 1, \cdots, s + t)$. Each of the arcs determines c_s as a focal point of P, since otherwise there would exist an $\epsilon > 0$ such that for every c on $c_s \leq c \leq c_s + \epsilon$ the index of λ^* as a characteristic value of $B(P: c)$ is greater than t. This, however, is impossible in view of the definition of c_s and since $\lambda_{s+t+1}(c)$ remains greater than λ^* for c in a neighborhood of c_s. Hence the index of c_s as a focal point of P relative to (6.4) for $\lambda = \lambda^*$ is equal to g.

Corresponding to a given problem B satisfying hypotheses (H), the normal boundary problem determined by the $I[\eta]$, $\Phi_a = 0$ belonging to B, together with the end conditions $\eta_i(a) = 0$, $\eta_i(b) = 0$, will be termed the *associated null end point problem.* Clearly, the associated null end point problem is a sub-problem of B of dimension at most $2n - p$. In view of Theorems 5.4 and 6.1 we have immediately:

THEOREM 6.2. *If for a given problem B satisfying hypotheses (H) the associated null end point problem is a sub-problem of B of dimension d, then for a given L the number of conjugate points of $x = a$ relative to the equations*

(6.4) of B for $\lambda = L$, and located on the open interval $a < x < b$, is at least $V_L - d$ and at most V_L.

7. A problem non-linear in the characteristic parameter. We shall now consider a boundary problem involving a parameter λ in such a manner that for every value of λ there is defined a self-adjoint integro-differential system of the sort treated in the preceding sections. It will be supposed that the functions M_{ij} and the coefficients of the quadratic forms ω, Q involve a real parameter λ, and are continuous in their arguments for $a \leq x,\ t \leq b$, $\mathfrak{C}: \mathfrak{C}_1 < \lambda < \mathfrak{C}_2$; the coefficients of Φ_α and Ψ_γ are supposed to be independent of λ. Moreover, it is supposed that for each value of λ on $\mathfrak{C}$ hypotheses (H) of Section 2 are satisfied. For brevity, these conditions will be denoted by (H^λ). The coefficients A_{ij}, B_{ij}, $l_{\alpha j}$, $m_{\alpha j}$ occurring in the solution (2.7) of the system (2.6) now depend upon λ, and we may write the canonical form of our integro-differential system as

$$\begin{aligned} &L_i[\eta,\zeta:\lambda] \equiv \eta'_i - A_{ij}(x:\lambda)\eta_j - B_{ij}(x:\lambda)\zeta_j = 0, \\ (7.1)\quad &L_{n+i}[\eta,\zeta:\lambda] \equiv \zeta'_i - C_{ij}(x:\lambda)\eta_j \\ &\qquad\qquad + A_{ji}(x:\lambda)\zeta_j - \int_a^b M_{ij}(x,t:\lambda)\eta_j(t)\,dt = 0, \\ &s_\sigma[\eta,\zeta:\lambda] = 0, \qquad\qquad (i,j = 1,\cdots,n;\ \sigma = 1,\cdots,2n). \end{aligned}$$

The boundary conditions $s_\sigma = 0$ have the explicit form (2.5′), where it is to be remembered that the coefficients of Q depend upon λ. A value λ will be said to be a characteristic value of this boundary problem if there exist functions $\eta_i(x)$, $\zeta_i(x)$ of class C^1, not all identically zero on ab and satisfying the system (7.1). The expression (1.1) for this new problem will be denoted by $I[\eta:\lambda]$.

In order to discuss the characteristic values of (7.1) we consider the auxiliary boundary problem

$$(7.2) \qquad L_i[\eta,\zeta:\lambda] = 0, \quad L_{n+i}[\eta,\zeta:\lambda] + \mu\eta_i = 0, \quad s_\sigma[\eta,\zeta:\lambda] = 0,$$

which involves the parameter μ linearly. In view of the above hypotheses and the results of Section 4, for each value of λ on $\mathfrak{C}$ there are infinitely many characteristic values $\mu_1(\lambda) \leq \mu_2(\lambda) \leq \cdots$ of (7.2). The corresponding characteristic solutions $\eta_i = y_{is}(x:\lambda)$, $\zeta_i = z_{is}(x;\lambda)$ are supposed orthonormal in the sense that

$$\int_a^b y_{is}(x:\lambda)y_{it}(x:\lambda)\,dx = \delta_{st} \qquad (s,t = 1,2,\cdots).$$

The class of admissible arcs η such that $K[\eta] = 1$, $K[y_\kappa(x:\lambda);\eta] = 0$

$(\kappa = 1, \cdots, s-1)$ will be denoted by $\mathfrak{S}_s(\lambda)$. It follows from Section 4 that $\mu_s(\lambda)$ is the minimum of $I[\eta:\lambda]$ in the class $\mathfrak{S}_s(\lambda)$. Clearly a value λ is a characteristic value of (7.1) if and only if there is an integer s such that $\mu_s(\lambda) = 0$; moreover, the index of λ as a characteristic value of (7.1) is equal to the index of $\mu = 0$ as a characteristic value of the corresponding system (7.2).

THEOREM 7.1. *If hypotheses* (H^λ) *are satisfied, the characteristic values* $\mu_s(\lambda)$ *of* (7.2) *are continuous functions of* λ *on* $\mathfrak{C}$.

The proof of this lemma uses Lemma 3.2. In view of the continuity of the coefficients of $I[\eta:\lambda]$ we have that for a given bounded and closed subinterval $\mathfrak{C}_0 : \lambda_0 \leq \lambda \leq \lambda'_0$ of $\mathfrak{C}$ there exist positive constants $k(\mathfrak{C}_0)$, $l(\mathfrak{C}_0)$, $m(\mathfrak{C}_0)$ such that for λ on $\mathfrak{C}_0$ the inequalities

$$(7.3)\qquad \int_a^b \eta'_i\eta'_i dx \leq m(\mathfrak{C}_0) I[\eta:\lambda] + l(\mathfrak{C}_0) K[\eta],$$

$$(7.4)\quad I[\eta:\lambda] \leq k(\mathfrak{C}_0)[\eta_i(a)\eta_i(a) + \eta_i(b)\eta_i(b) + \int_a^b (\eta'_i\eta'_i + \eta_i\eta_i)\,dx],$$

hold for arbitrary differentially admissible arcs η. In view of (7.4) and Theorem 5.1 there are constants $l_s(\mathfrak{C}_0)$ $(s = 1, 2, \cdots)$ such that $\mu_s(\lambda) \leq l_s(\mathfrak{C}_0)$ $(s = 1, 2, \cdots)$ for λ on $\mathfrak{C}_0$.

Now let λ, λ' be values on $\mathfrak{C}_0$ and choose constants $d_1, \cdots, d_s$ such that the arc $u_i = y_{i1}(x:\lambda')d_1 + \cdots + y_{is}(x:\lambda')d_s$ belongs to $\mathfrak{S}_s(\lambda)$. We then have $\mu_s(\lambda) \leq I[u:\lambda]$, $I[u:\lambda'] \leq \mu_s(\lambda')$. Hence

$$\begin{aligned}\mu_s(\lambda) - \mu_s(\lambda') \leq I[u:\lambda] - I[u:\lambda'] &\leq \epsilon_1(\lambda,\lambda')[u_i(a)u_i(a) + u_i(b)u_i(b)] \\ &\quad + \epsilon_2(\lambda,\lambda')\int_a^b (u'_iu'_i + u_iu_i)\,dx,\end{aligned}$$

where $\epsilon_1(\lambda,\lambda')$, $\epsilon_2(\lambda,\lambda')$ are independent of the particular arc u, are symmetric in (λ,λ'), and tend to zero with $|\lambda - \lambda'|$ uniformly for λ, λ' on $\mathfrak{C}_0$. This inequality is a consequence of the continuity of the coefficients of $I[\eta:\lambda]$ as functions of λ. Let $\phi(x) = 2(x-a)/(b-a) - 1$. Then $\phi(a) = -1$, $\phi(b) = 1$, $|\phi'| = 2/(b-a)$, and $|\phi| \leq 1$ on ab. Now

$$\begin{aligned}|u_i(a)u_i(a) + u_i(b)u_i(b)| &\leq \int_a^b |2\phi u_iu'_i + \phi' u_iu_i|\,dx, \\ &\leq \int_a^b [|\phi|\,u'_iu'_i + (|\phi| + |\phi'|)u_iu_i]\,dx, \\ &\leq \int_a^b u'_iu'_i dx + M_1,\end{aligned}$$

where $M_1 = 1 + 2/(b-a)$. Hence, for $\epsilon_3 = \epsilon_1 + \epsilon_2$, $\epsilon_4 = M_1\epsilon_1 + \epsilon_2$ we have,

$$(7.5)\quad \begin{aligned} \mu_s(\lambda) - \mu_s(\lambda') &\leq \epsilon_3(\lambda, \lambda') \int_a^b u'_i u'_i dx + \epsilon_4(\lambda, \lambda'), \\ &\leq \epsilon_3(\lambda, \lambda')[m(\mathfrak{C}_0) I[u:\lambda'] + l(\mathfrak{C}_0)] + \epsilon_4(\lambda, \lambda'), \\ &\leq \epsilon_3(\lambda, \lambda')[m(\mathfrak{C}_0)\, l_s(\mathfrak{C}_0) + l(\mathfrak{C}_0)] + \epsilon_4(\lambda, \lambda'), \end{aligned}$$

since $I[u:\lambda'] \leq \mu_s(\lambda') \leq l_s(\mathfrak{C}_0)$. As ϵ_3, ϵ_4 are symmetric in (λ, λ'), it follows that $|\mu_s(\lambda) - \mu_s(\lambda')|$ does not exceed the right-hand member of inequality (7.5), and the continuity of $\mu_s(\lambda)$ on $\mathfrak{C}_0$ is immediate.

We shall now consider a system which satisfies the following additional hypotheses:

$(H_4{}^\lambda)$ For λ sufficiently near $\mathfrak{C}_1$, $I[\eta:\lambda] > 0$ for all non-identically vanishing admissible arcs η.

$(H_5{}^\lambda)$ For each integer s, if $\mu_s(\lambda) = 0$ for $\lambda = \lambda'$, then $\mu_s(\lambda) < 0$ for $\lambda > \lambda'$.

Suppose that B is the above defined problem involving $\omega(x, \eta, \eta':\lambda)$, $Q[\eta(a), \eta(b):\lambda]$, $M_{ij}(x, t:\lambda)$, $\phi_\alpha(x, \eta, \eta')$, $\Psi_\gamma[\eta(a), \eta(b)]$, that B^* is a second problem involving $\omega^*(x, \eta, \eta':\lambda)$, $Q^*[\eta(a), \eta(b):\lambda]$, $M^*_{ij}(x, t:\lambda)$, $\Phi_\alpha(x, \eta, \eta')$, $\Psi^*_\nu[\eta(a), \eta(b)]$, and that B and B^* each satisfy hypotheses (H^λ), $(H_4{}^\lambda)$, $(H_5{}^\lambda)$. Let $N(L)$ and $P(L)$ denote, respectively, the number of negative and positive zeros of $D(\rho:L)$ defined by (5.1) in terms of the difference end-form $d[\eta(a), \eta(b):L] = Q^*[\eta(a), \eta(b):L] - Q[\eta(a), \eta(b):L]$. Similarly, the N_0, P_0 occurring in Theorem 5.5 are now replaced by $N_0(L)$, $P_0(L)$. By considering the curves $\mu = \mu_s(\lambda)$ it is readily seen that the conclusions of Theorem 5.1, Theorem 5.4 and its corollary, together with the inequalities involving V_L, W_L, V^*_L, W^*_L occurring in the corollaries to Theorems 5.3 and 5.5, hold for the above defined problems B and B^*, with the understanding that the $I[\eta]$, $I^*[\eta]$, N, P, N_0, P_0 occurring in the statements of these results are now replaced by $I[\eta:\lambda]$, $I^*[\eta:\lambda]$, $N(L)$, $P(L)$, $N_0(L)$, $P_0(L)$. Corresponding to the corollary of Theorem 5.2 we now have under the hypotheses corresponding to those of Theorem 5.2 that if λ_s exists for B and if h_s denote the number of characteristic values of $D(B, B^*)$ less than λ_s, then $\lambda^*_{s+h_s}$, if it exists, is not less than λ_s. It is also readily seen that Theorems 6.1 and 6.2 hold for a problem B satisfying the above hypotheses. It is to be noted that for the above theorems we have not supposed that B and B^* have infinitely many characteristic values.

Applying the results of Theorems 4.2 and 4.3 to the problem (7.2), we see that $(H_5{}^\lambda)$ holds if $I[\eta:\lambda]$ is a proper monotone decreasing function of λ for all non-identically vanishing admissible arcs η. This latter condition is in turn assured if the coefficients of the functions occurring in $I[\eta:\lambda]$ have suitable monotone properties as functions of λ. Such conditions are assumed in the classical Sturmian theory for a second-order differential equation, and have also been used by Morse [16].

We shall now consider a problem (7.1) which satisfies the following condition.

(H_6^λ) There is a function $g(\lambda', \lambda)$ defined for $\mathfrak{C}_1 < \lambda', \lambda < \mathfrak{C}_2$, $\lambda > \lambda'$ which is positive and such that if $\lambda > \lambda'$ then

$$I[\eta:\lambda] < g(\lambda', \lambda) I[\eta:\lambda']$$

for all non-identically vanishing admissible arcs η.

THEOREM 7.2. *If a problem satisfies hypotheses* (H^λ), (H_4^λ), (H_6^λ), *then* (H_5^λ) *is also satisfied.*

For suppose that for a value λ' there is an integer s such that $\mu_s(\lambda') = 0$. For arbitrary constants d_κ $(\kappa = 1, \cdots, s)$ the arc $\eta_i = \sum_{\kappa=1}^{s} y_{i\kappa}(x:\lambda') d_\kappa$ is such that $I[\eta:\lambda'] \leq \mu_s(\lambda') = 0$. By ($H_6^\lambda$), for $\lambda > \lambda'$ each of these arcs is such that $I[\eta:\lambda] < g(\lambda', \lambda) I[\eta:\lambda'] \leq 0$. It then follows from the extremizing properties of the characteristic values of (7.2) that $\mu_s(\lambda) < 0$ for $\lambda > \lambda'$.

Now consider a problem of the form formulated in Section 2 which satisfies in addition to hypotheses (H) the condition that $I[\eta] > 0$ for all non-identically vanishing admissible arcs λ. Let $\tilde{B}$ denote the boundary problem determined by

$$(7.6) \qquad I[\eta:\lambda] = I[\eta] - \lambda \left\{ \mathfrak{Q}[\eta(a), \eta(b)] + \int_a^b K_{ij}(x) \eta_i \eta_j dx + \int_a^b \int_a^b N_{ij}(x, t) \eta_i(x) \eta_j(t) \, dx dt \right\},$$

together with the differential equations $\Phi_\alpha = 0$ and the end-conditions $\Psi_\gamma = 0$ belonging to the first problem. It is to be understood that $\mathfrak{Q}$ is a quadratic form in $[\eta_i(a), \eta_i(b)]$, and that $\| K_{ij}(x) \|$, $\| N_{ij}(x, t) \|$ are matrices whose elements are continuous with $K_{ij}(x) = K_{ji}(x)$, $N_{ij}(x, t) = N_{ji}(t, x)$. Clearly this problem $\tilde{B}$ satisfies (H^λ) and (H_4^λ) on $0 < \lambda < +\infty$. It also satisfies (H_6^λ) since for $\lambda > \lambda' > 0$ and arbitrary admissible arcs η,

$$I[\eta:\lambda] - I[\eta:\lambda'] = [(\lambda - \lambda')/\lambda'] \{ I[\eta:\lambda'] - I[\eta] \}.$$

Using the notation of Section 2, this system may be written as

$$(7.7) \qquad \begin{aligned} & L_i[\eta, \zeta] = 0, \quad L_{n+i}[\eta, \zeta | 0] + \lambda \left\{ K_{ij}(x) \eta_j + \int_a^b N_{ij}(x, t) \eta_j(t) \, dt \right\} = 0, \\ & s_\sigma[\eta, \zeta : \lambda] = 0, \end{aligned}$$

where the boundary conditions are linear in λ. The above results give oscillation and comparison theorems for the positive characteristic values and associated characteristic solutions of (7.6). To obtain corresponding results for the negative characteristic values, one need only replace the form $\mathfrak{Q}$ and the functions K_{ij}, N_{ij} by their negatives, and consider the resulting system, again for $0 < \lambda < +\infty$. In case the functions $M_{ij}(x, t)$ occurring in $I[\eta]$ and the functions $N_{ij}(x, t)$ are identically zero, system (7.7) reduces to the differential system considered by the author [18].

We shall now consider the question of the existence of infinitely many characteristic values of (7.1). Clearly such a system which satisfies (H^λ), $(H_4{}^\lambda)$, $(H_5{}^\lambda)$ will have infinitely many characteristic values if and only if for each integer s there is a value λ' such that $\mu_s(\lambda') = 0$. In view of the extremizing properties of the values $\mu_s(\lambda)$ we have that a system (7.1) satisfying the above hypotheses will have at least k characteristic values if and only if there are k admissible functions $\eta_{i\kappa}(x)$ $(\kappa = 1, \cdots, s)$ and a value λ' such that for arbitrary constants $(d_\kappa) \neq (0_\kappa)$ we have $I[\eta_\kappa d_\kappa : \lambda'] < 0$. In view of the analogue of the corollary to Theorem 5.5, an arbitrary problem (7.1) satisfying the above hypotheses will have an infinity of characteristic values if and only if there are infinitely many characteristic values of the normal problem determined by the null end-conditions $\eta_i(a) = 0 = \eta_i(b)$, and having $M_{ij}(x, t : \lambda)$, $\omega(x, \eta, \eta' : \lambda)$, $\Phi_\alpha(x, \eta, \eta')$ in common with the given problem. The following special criterion is readily proved.

THEOREM 7.3. *Suppose that for a problem* (7.1) *satisfying* (H^λ), $(H_4{}^\lambda)$, $(H_5{}^\lambda)$ *there are functions* $R(\lambda) > 0$, $P(\lambda)$ *on* $\mathfrak{C}$ *such that* $P(\lambda) \to +\infty$ *as* $\lambda \to \mathfrak{C}_2$, *and for arbitrary admissible arcs* η *vanishing at* $x = a$ *and* $x = b$ *we have*

$$\int_a^b 2\omega(x, \eta, \eta' : \lambda)\,dx + \int_a^b \int_a^b M_{ij}(x, t : \lambda)\eta_i(x)\eta_j(t)\,dx\,dt \leq \int_a^b R(\lambda)[\eta'_i\eta'_i - P(\lambda)\eta_i\eta_i]\,dx. \tag{7.8}$$

Then this problem has infinitely many real characteristic values on $\mathfrak{C}$.

It is to be noted that (7.8) is assured if there are functions $R(\lambda) > 0$, $P_1(\lambda)$, $P_2(\lambda)$ such that

$$2\omega(x, \eta, \pi : \lambda) \leq R(\lambda)[\pi_i\pi_i - P_2(\lambda)\eta_i\eta_i] \tag{7.9}$$

for arbitrary sets (η_i, π_i), the inequality

$$\int_a^b \int_a^b M_{ij}(x, t : \lambda)\eta_i(x)\eta_j(t)\,dx\,dt \leq R(\lambda)P_1(\lambda) \int_a^b \eta_i\eta_i\,dx \tag{7.10}$$

holds for arbitrary continuous functions $\eta_i(x)$, and $P_1(\lambda) - P_2(\lambda) \to -\infty$ as $\lambda \to \mathfrak{C}_2$. Using the Schwarz integral inequality, it follows that (7.10) is certainly true if there is a function $P_3(\lambda) \geqq 0$ such that $|M_{ij}(x, t:\lambda)| \leqq R(\lambda)P_3(\lambda)/[n(b-a)]$ for $a \leqq x$, $t \leqq b$, λ on $\mathfrak{C}$, and $P_3(\lambda) - P_2(\lambda) \to \infty$ as $\lambda \to \mathfrak{C}_2$.

In conclusion, we shall consider the question of infinitely many characteristic values of the problem $\tilde{B}$ given by (7.7), where it is understood that the quantities $I[\eta]$, Q, Φ_α, Ψ_γ involved satisfy hypotheses (H) of Section 2, and $I[\eta] > 0$ for all non-identically vanishing admissible arcs η. For brevity, we write

$$(7.11) \qquad \mathfrak{N}[\eta] \equiv \int_a^b \int_a^b N_{ij}(x, t)\eta_i(x)\eta_j(t)\,dxdt,$$

$$\mathfrak{K}[\eta] = \mathfrak{Q}[\eta(a), \eta(b)] + \int_a^b K_{ij}\eta_i\eta_j dx + \mathfrak{N}[\eta].$$

The following theorem may be proved by the method used by the author ([19], pp. 846-848) in considering the corresponding result for a differential system.

THEOREM 7.4. *Suppose the problem $\tilde{B}$ satisfies the additional condition that $\mathfrak{K}[\eta] > 0$ (< 0) for arbitrary non-identically vanishing admissible arcs η. Then the characteristic values of $\tilde{B}$ are all positive (negative), and are infinite in number.*

THEOREM 7.5. *Suppose that the quantities $I[\eta]$, Q, Φ_α, Ψ_γ involved in the problem $\tilde{B}$ satisfy hypotheses (H), but not necessarily the condition that $I[\eta] > 0$ for all non-identically admissible arcs η. If, however, we also have:* (i) *the matrix $\| K_{ij} \|$ is positive definite on ab;* (ii) $\mathfrak{Q}[\eta(a), \eta(b)] + \mathfrak{N}[\eta] \geqq 0$ *for arbitrary admissible arcs η, then $\tilde{B}$ has infinitely many positive and only a finite number of negative characteristic values.*

Since $\| K_{ij} \|$ is positive definite, we have by the Schwarz inequality that there is a constant l such that $|M[\eta]| \leqq l \int_a^b K_{ij}\eta_i\eta_j dx$. It follows (see Morse [15], p. 533, and Reid [18], p. 786) that there is a positive constant λ_0 such that

$$0 < I[\eta] + \lambda_0 \int_a^b K_{ij}\eta_i\eta_j dx \leqq I[\eta] + \lambda_0 \mathfrak{K}[\eta]$$

for all non-identically vanishing admissible arcs η. If $I[\eta]$ is replaced by $I[\eta] + \lambda_0\mathfrak{K}[\eta]$, the modified problem is equivalent to the original problem by a linear change of parameter. By Theorem 7.4, however, the characteristic values of the modified problem are all positive and infinite in number. Hence the original problem has infinitely many characteristic values of which only a finite number are negative.

One may also prove for $\tilde{B}$ a theorem analogous to Theorem 6.1 of Reid [18], but such a theorem will not be explicitly stated here.

8. Properties of integro-differential systems. In this section we shall state some significant properties of a general integro-differential system of the first order.

$$(8.1)\quad \begin{aligned} L_i[y] &\equiv y'_i - A_{ij}(x)y_j - \int_a^b C_{ij}(x,t)y_j(t)\,dt = 0, \\ s_i[y] &\equiv M_{ij}y_j(a) + N_{ij}y_j(b) = 0 \qquad (i, j = 1, \cdots, m). \end{aligned}$$

For brevity no proofs will be given, since these results may be established by a combination of the fundamental theorems of the Fredholm integral equation theory and the methods used by Bliss [2] in treating a differential boundary problem. It will be assumed that the functions $A_{ij}(x)$, $C_{ij}(x,t)$ are continuous on $a \leq x,\ t \leq b$, and that the constant matrix $\| M_{ij}; N_{ij} \|$ has rank m. Throughout this section the subscripts i, j, k, l will have the range $1, \cdots, m$. It is to be noted that the integro-differential system of Section 2 is of the form (8.1) for $m = 2n$ and $y \equiv [\eta_1, \cdots, \eta_n, \zeta_1, \cdots, \zeta_n]$.

Let $p_k = P_{kj}$, $q_k = Q_{kj}$ $(j = 1, \cdots, m)$ be linearly independent solutions of the equations $M_{ik}p_k - N_{ik}q_k = 0$ $(i = 1, \cdots, m)$. In accordance with the definition of Bliss [2] for a differential system, we say that the boundary conditions $t_i[z] \equiv z_j(a)P_{ji} + z_j(b)Q_{ji}$ $(i = 1, \cdots, m)$ are adjoint to the boundary conditions $s_i[y] = 0$. Furthermore, the integro-differential system

$$(8.2)\quad M_i[z] \equiv z'_i + z_jA_{ji}(x) + \int_a^b z_j(t)C_{ji}(t,x)\,dt = 0, \quad t_i[z] = 0$$

will be said to be the system adjoint to (8.1). It is to be noted that if $y_i(x)$ and $z_i(x)$ are solutions of $L_i[y] = 0$ and $M_i[z] = 0$, respectively, then $y_i(x)z_i(x)$ is constant on ab.

Let $\| M^*_{ij} \|$, $\| N^*_{ij} \|$, $\| P^*_{ij} \|$, $\| Q^*_{ij} \|$ be matrices such that the matrices of $2m$ rows and columns

$$\left\| \begin{matrix} M_{ij} & N_{ij} \\ M^*_{ij} & N^*_{ij} \end{matrix} \right\|, \qquad \left\| \begin{matrix} -P^*_{ij} & -P_{ij} \\ Q^*_{ij} & Q_{ij} \end{matrix} \right\|$$

are reciprocals. Then the expressions

$$s^*_i[y] \equiv M^*_{ij}y_j(a) + N^*_{ij}y_j(b), \quad t^*_i[z] \equiv z_j(a)P^*_{ji} + z_j(b)Q^*_{ji}$$

are such that for arbitrary values of $y_i(x)$, $z_i(x)$ at $x = a$ and $x = b$ we have (Bliss [2], p. 565)

$$t^*_i[z]s_i[y] + t_i[z]s^*_i[y] = z_i(x)y_i(x)\big|_a^b. \tag{8.3}$$

Now let $B_{ij}(x)$ be arbitrary continuous functions such that the differential system

$$y'_i - B_{ij}(x)y_j = 0, \quad s_i[y] = 0 \qquad (i, j = 1, \cdots, m) \tag{8.4}$$

is incompatible. Such functions may be chosen in infinitely many ways. Let $y_i = Y_{ij}(x)$ $(j = 1, \cdots, m)$ be a fundamental set of solutions of the differential equations of (8.4), and set $\| D_{ij} \| = \| s_i(Y_j) \|$. We shall denote by $\| G_{ij}(x, t) \|$ the Green's matrix for the incompatible system (8.4).

If $f_i(x)$ are arbitrary continuous functions and h_i are given constants, then every solution $y_i = y_i(x)$ $(i = 1, \cdots, m)$ of the non-homogeneous system

$$L_i[y] = f_i(x), \quad s_i[y] = h_i \qquad (i = 1, \cdots, m) \tag{8.5}$$

is also a solution of the integral system

$$y_i(x) = \int_a^b K_{ij}(x, t)y_j(t)\,dt + F_i(x), \tag{8.6}$$

where

$$K_{ij}(x, t) = G_{ik}(x, t)[A_{kj}(t) - B_{kj}(t)] + \int_a^b G_{ik}(x, s)C_{kj}(s, t)\,ds,$$

$$F_i(x) = \int_a^b G_{ij}(x, t)f_j(t)\,dt + Y_{ik}(x)D^{-1}{}_{kj}h_j,$$

and conversely. Suppose that the system (8.1) has only the identically vanishing solution. Then the homogeneous integral system

$$y_i(x) \equiv \int_a^b K_{ij}(x, t)y_j(t)\,dt \tag{8.7}$$

has only the trivial solution $y_i(x) \equiv 0$, and there exists a resolvent matrix $\| R_{ij}(x, t) \|$ for the system (8.6). It may be shown that if (8.1) is incompatible, then for arbitrary continuous functions $f_i(x)$ and constants h_i the system (8.5) has a unique solution given by

$$y_i(x) = \int_a^b \mathfrak{g}_{ij}(x, t)f_j(t)\,dt + c_{ij}(x)h_j, \tag{8.8}$$

where

$$
(8.9)\qquad \begin{aligned}
\mathcal{G}_{ij}(x,t) &= G_{ij}(x,t) + \int_a^b R_{ik}(x,s)G_{kj}(s,t)\,ds,\\
c_{ij}(x) &= [Y_{il}(x) + \int_a^b R_{ik}(x,s)Y_{kl}(s)\,ds]D^{-1}{}_{lj}.
\end{aligned}
$$

Corresponding to the notation of Tamarkin [21] and Jonah [11], the matrix $\| \mathcal{G}_{ij}(x,t) \|$ will be called a Green's matrix for the integro-differential boundary problem (8.1). It follows readily that for an incompatible system (8.1) the Green's matrix is unique, and hence independent of the particular choice of the matrix $\| B_{ij} \|$.

Using the properties of the ordinary Green's matrix for the differential system (3.4), together with results of the Fredholm integral equation theory, one may prove the following results.

THEOREM 8.1. *The number of linearly independent solutions of the homogeneous system* (8.1) *is the same as the number of such solutions of the adjoint system* (8.2).

THEOREM 8.2. *If the integro-differential systems* (8.1) *and* (8.2) *are incompatible, and* $\| \mathcal{G}_{ij}(x,t) \|$ *and* $\| \mathcal{H}_{ij}(x,t) \|$ *are the respective Green's matrices, then* $\mathcal{G}_{ij}(x,t) = -\mathcal{H}_{ji}(t,x)$.

THEOREM 8.3. *If the system* (8.1) *has exactly r linearly independent solutions* $y_i = y_{i\nu}(x)$ $(\nu = 1,\cdots,r)$, *then the non-homogeneous system* (8.5) *has a solution if and only if the condition*

$$(8.10)\qquad \int_a^b z_i(x)f_i(x)\,dx = h_i t^*{}_i[z]$$

is satisfied by every solution $z_i(x)$ *of the adjoint system* (8.2); *moreover, the most general solution of* (8.5) *is of the form*

$$y_i = y^*{}_i(x) + y_{i\nu}(x)c_\nu,$$

where $y^*{}_i(x)$ *is a particular solution and the* c_ν*'s are arbitrary constants.*

If the system (8.1) is compatible one may define a generalized Green's matrix for this system, using methods similar to those employed by the author [17] in considering compatible differential systems. Since, however, no use is made of generalized Green's matrices in the consideration of the problem defined in Section 2, the explicit form of the generalized Green's matrix will not be given.

A boundary problem (8.1) will be said to be self-adjoint if there exists a non-singular matrix $\| T_{ij}(x) \|$ whose elements are of class C^1 on $a \leq x \leq b$, and satisfying the conditions

$$
\begin{aligned}
&T_{ik}A_{kj} + A_{ki}T_{kj} + T'_{ij} \equiv 0, \quad T_{ik}(x)C_{kj}(x,t) + T_{kj}(t)C_{ki}(t,x) \equiv 0 \\
&M_{ik}T^{-1}{}_{kl}(a)M_{jl} - N_{ik}T^{-1}{}_{kl}(b)N_{jl} = 0 \qquad (i, j = 1, \cdots, m).
\end{aligned}
\tag{8.11}
$$

This definition of self-adjointness corresponds to that given by Bliss ([2], p. 569) for a differential boundary problem. If a matrix $\| T_{ij}(x) \|$ satisfies the above conditions it follows readily that for every solution $y_i(x)$ of (8.1) there is a solution $z_i(x)$ of (8.2) given by $z_i(x) = T_{ij}(x)y_j(x)$, and conversely. In a manner similar to that used by Bliss [2] to prove the corresponding result for a differential boundary problem, one may establish the following theorem.

THEOREM 8.4. *Suppose the problem* (8.1) *is incompatible and is self-adjoint with respect to the matrix* $\| T_{ij}(x) \|$. *Then the functions of the Green's matrix* $\| \mathfrak{G}_{ij}(x,t) \|$ *satisfy the relations*

$$
T_{ik}(x)\mathfrak{G}_{kj}(x,t) + T_{kj}(t)\mathfrak{G}_{ki}(t,x) \equiv 0. \tag{8.12}
$$

As indicated in Section 3, the integro-differential system (2.4′), (2.5′) is self-adjoint with respect to the constant matrix (3.6).

BIBLIOGRAPHY.

1. Birkhoff and Hestenes, "Natural isoperimetric conditions in the calculus of variations," *Duke Mathematical Journal*, vol. 1 (1935), pp. 198-286.

2. Bliss, "A boundary value problem for a system of ordinary linear differential equations of the first order," *Transactions of the American Mathematical Society*, vol. 28 (1926), pp. 561-589.

3. Bliss, "The problem of Lagrange in the calculus of variations," *American Journal of Mathematics*, vol. 52 (1930), pp. 673-744.

4. Bliss, *The Problem of Bolza in the Calculus of Variations*, mimeographed notes of lectures delivered at the University of Chicago, Winter Quarter, 1935.

5. Caratheodory, *Variationsrechnung*, Berlin, 1934.

6. Courant, "Über die Anwendung der Variationsrechnung in der Theorie der Eigenschwingungen und über neue Klasse von Funktionalgleichungen," *Acta Mathematica*, vol. 49 (1927), pp. 1-68.

7. Courant-Hilbert, *Methoden der Mathematischen Physik*, I, Berlin, 1924.

8. Hölder, "Die Lichtensteinsche Methode für die Entwicklung der zweiten Variation, angewandt auf das Problem von Lagrange," *Prace Matematyczno-Fizyczne*, vol. 42 (1935), pp. 307-346.

9. Hu, "The problem of Bolza and its accessory boundary value problem," *Contributions to the Calculus of Variations 1931-32*, The University of Chicago Press, pp. 361-443.

10. Ince, *Ordinary Differential Equations*, London, 1927.

11. Jonah, "The Green's matrix and expansion problem for systems of integro-differential equations," *Dissertation* [Brown University, 1930]. For abstract, see *Bulletin of the American Mathematical Society*, vol. 36 (1930), p. 184.

12. Lichtenstein, "Über eine Integro-Differentialgleichung und die Entwicklung willkürlicher Funktionen nach der Eigenfunktionen," *Schwarz' Festschrift*, Berlin, 1924, pp. 274-285.

13. Lichtenstein, "Zur Variationsrechnung, II," *Journal für die reine und angewandte Mathematik*, vol. 164 (1931), pp. 194-216.

14. Morse, "A generalization of the Sturm separation and comparison theorems in n-space," *Mathematische Annalen*, vol. 103 (1930), pp. 72-91.

15. Morse, "Sufficient conditions in the problem of Lagrange with variable end conditions," *American Journal of Mathematics*, vol. 53 (1931), pp. 517-546.

16. Morse, "The calculus of variations in the large," *American Mathematical Colloquium Publications*, vol. 18 (1934).

17. Reid, "Generalized Green's matrices for compatible systems of differential equations," *American Journal of Mathematics*, vol. 54 (1932), pp. 443-459.

18. Reid, "A boundary value problem associated with the calculus of variations," *American Journal of Mathematics*, vol. 54 (1932), pp. 769-790.

19. Reid, "Analogues of the Jacobi condition for the problem of Mayer in the calculus of variations," *Annals of Mathematics*, vol. 35 (1934), pp. 836-848.

20. Reid, "The theory of the second variation for the non-parametric problem of Bolza," *American Journal of Mathematics*, vol. 57 (1935), pp. 573-586.

21. Tamarkin, "The notion of the Green's matrix in the theory of integro-differential equations, I and II," *Transactions of the American Mathematical Society*, vol. 29 (1927), pp. 755-800 and vol. 32 (1930), pp. 860-868.

22. Wiggin, "A boundary value problem of the calculus of variations," *Contributions to the Calculus of Variations* 1933-37, The University of Chicago Press, pp. 245-275.

THE UNIVERSITY OF CHICAGO,
CHICAGO, ILLINOIS.

A PARTIAL DIFFERENTIAL EQUATION ASSOCIATED WITH POISSON'S WORK ON THE THEORY OF SOUND.*

By H. BATEMAN.

Introduction. In his famous memoir of 1808 on the theory of sound, in which he discussed the theory of sound waves of finite amplitude, Poisson[1] also made some important advances in the theory of sound waves of small amplitude. In particular he attacked the problem of solving the equation of wave-propagation in three dimensions by using the mean value of a function over a sphere; a method which eventually led to a general solution and proved very fruitful in potential theory. This method has been supplemented by the consideration of mean values around circles on a sphere.[2]

To determine the velocity components of the individual particles of air Poisson tried to solve the wave equation by means of an infinite series of powers of the inverse distance in which the n-th coefficient is an integral of order n. The integrands of the various integrals are connected by a recurrence relation of the first order involving partial derivatives of the first two orders. Poisson found that these integrands could be obtained from a generating function which in turn satisfies a certain partial differential equation, namely, the one designated as equation (1) below. Incidentally, this differential equation also occurs in the above mentioned theory of mean values around circles on a sphere.

Here it is pointed out that the same partial differential equation may be derived from the wave equation by a simple transformation suggested by Poisson's work. The general problem of obtaining a wave function from a solution of (1) is considered. It is found that $(1/r)U(w, \mu, \phi)$, where $\mu = \cos\theta$ and r, θ, ϕ are polar coördinates, is a wave function provided w is defined by equation (6).

Equation (1) has particular solutions represented by products of Legendre functions. By comparing these solutions with another solution of the wave equation expansions involving Legendre functions are suggested. The determi-

---

* Received December 6, 1937.

[1] S. D. Poisson, *Bull. Soc. Philom.*, vol. 1 (1807), p. 19; *Journal de l'École Polytechnique*, t. 7, Cah. 14 (1808), pp. 319-392.

[2] H. Bateman, *Proceedings of the National Academy of Sciences*, vol. 16 (1930), pp. 205-211; *Annals of Mathematics* (2), vol. 31 (1930), pp. 158-162.

nation of the coefficients in these expansions requires the evaluation of a class of integrals discussed in an accompanying paper.

Relation to the wave equation. The differential equation which forms the subject of this paper is

$$(1)\qquad \frac{\partial}{\partial w}\left[(1-w^2)\frac{\partial U}{\partial w}\right]=\frac{\partial}{\partial \mu}(1-\mu^2)\frac{\partial U}{\partial \mu}+\frac{1}{1-\mu^2}\frac{\partial^2 U}{\partial \phi^2}.$$

Poisson's work indicates that the wave equation

$$(2)\qquad \frac{\partial^2 W}{\partial t^2}=\frac{\partial^2 W}{\partial x^2}+\frac{\partial^2 W}{\partial y^2}+\frac{\partial^2 W}{\partial z^2}$$

has particular solutions of the type

$$(3)\qquad W=\frac{1}{r}U(w,\mu,\phi)$$

where U is a solution of (1), $\mu=\cos\theta$ and r, θ, ϕ are polar coördinates, and w is defined by

$$(4)\qquad w=\frac{t}{r}.$$

This may be verified by transforming the wave equation by the substitution $z=\mu r$, $r^2=x^2+y^2+z^2$, $y=x\tan\phi$ after which it takes the form

$$(5)\quad \frac{\partial^2(rW)}{\partial t^2}=\frac{\partial^2(rW)}{\partial r^2}+\frac{1}{r^2}\left[\frac{\partial}{\partial \mu}(1-\mu^2)\frac{\partial}{\partial \mu}(rW)\right]+\frac{1}{r^2(1-\mu^2)}\frac{\partial^2}{\partial \phi^2}(rW).$$

Direct substitution of (3) and (4) in (5) then results in equation (1).

When we consider the general problem of finding a function $w=w(r,t)$ such that $W=(1/r)U(w,\mu,\phi)$ satisfies equation (5), and hence is a wave function, we are led to the two equations

$$\frac{\partial^2 w}{\partial t^2}=\frac{\partial^2 w}{\partial r^2}-\frac{2w}{r^2},\qquad \left(\frac{\partial w}{\partial t}\right)^2-\left(\frac{\partial w}{\partial r}\right)^2=\frac{1-w^2}{r^2}.$$

If we seek a solution of the type $w=\sum_{n=0}^{\infty}A_n(r)t^n$ we find that

$$(6)\qquad rw=C(r^2-t^2)+D+t(1-4CD)^{\frac{1}{2}}$$

where C and D are arbitrary constants. When $C=D=0$, Poisson's expression $w=t/r$ is obtained, while if $C=1/2a$, $D=a/2$ the value[3]

---

[3] H. Bateman, *loc. cit.*

$$w = \frac{r^2 - t^2 + a^2}{2ar} \tag{7}$$

is obtained.

An expression suggested by a solution of equation (1). The differential equation (1) has solutions of the type

$$\begin{aligned} U &= P_n(w) P_n^m(\mu) e^{im\phi} \\ U &= P_n(w) Q_n^m(\mu) e^{im\phi} \end{aligned} \tag{8}$$

where m and n are arbitrary constants. We shall take them to be non-negative integers. When we replace μ by $\tanh v$, the corresponding solutions of the wave equation are

$$\begin{aligned} W &= \frac{1}{r} P_n(w) P_n^m(\tanh v) e^{im\phi} \\ W &= \frac{1}{r} P_n(w) Q_n^m(\tanh v) e^{im\phi}. \end{aligned} \tag{9}$$

Another solution may be obtained directly from the wave equation by assuming that W is independent of z. Upon setting $Z = t$, $X = ix$, $Y = iy$ the wave equation reduces to Laplace's equation which is known to be satisfied by

$$W = \frac{1}{R} F\left(\frac{Z + R}{X + iY}\right), \qquad R^2 = X^2 + Y^2 + Z^2.$$

The values of $X + iY$ and R^2 expressed in terms of r, w, v and ϕ, where we now take $w = t/r$, are

$$\begin{aligned} X + iY &= ir \operatorname{sech} v e^{i\phi} \\ R^2 &= t^2 - x^2 - y^2 = r^2(w^2 - \operatorname{sech}^2 v) \end{aligned}$$

and hence the expression above for W assumes the form

$$W = \frac{1}{r\sqrt{w^2 - \operatorname{sech}^2 v}} f\left[e^{-i\phi}\{w \, ch \, v + (w^2 ch^2 v - 1)^{\frac{1}{2}}\} \right]$$

where $F(-is) = f(s)$. The argument of the function f may be simplified by writing $ch \, u = w \, ch \, v$. Then

$$W = \frac{1}{r\sqrt{w^2 - \operatorname{sech}^2 v}} f(e^{-i\phi + u}).$$

The conjugate complex of this expression is also a solution of the wave equation.

$$W = \frac{1}{r\sqrt{w^2 - \operatorname{sech}^2 v}}\, f_1(e^{i\phi+u}).$$

Taking $f_1(s) = s^m/2$ and $f(s) = s^{-m}/2$ and adding shows that

$$W = \frac{e^{im\phi}}{r\sqrt{w^2 - \operatorname{sech}^2 v}}\, ch\, mu = \frac{e^{im\phi}T_m(w\, ch\, v)}{r\sqrt{w^2 - \operatorname{sech}^2 v}} \tag{10}$$

is also a solution. $T_m(z)$ is Tchebycheff's polynomial defined as

$$T_m(z) = ch(m\, ch^{-1} z).$$

The solution (10) may be expanded in terms of the simple solutions (9). For example, when m is even, say $m = 2k$,

$$\sum_{n=0}^{\infty} (4n+1) P_{2n}(w) Q_{2n}^{2k}(\tanh v)/Q_{2n}^{2k+1}(0) = -\frac{T_{2k}(w\, ch\, v)}{r\sqrt{w^2 - \operatorname{sech}^2 v}} \text{ or } 0,$$
$$-1 < w < 1.$$

the first or second value on the right being taken accordingly as $w^2 - \operatorname{sech}^2 v$ is positive or negative. The determination of the cofficients in this series requires the evaluation of the integral

$$\int_0^v du\, P_{2n}\left(\frac{ch\, u}{ch\, v}\right) ch(2ku)\, du$$

which is discussed in an accompanying paper.

CALIFORNIA INSTITUTE OF TECHNOLOGY.

INTEGRALS INVOLVING LEGENDRE FUNCTIONS.*

By H. BATEMAN and S. O. RICE.

The preceding paper by one of us investigates the partial differential equation

$$\frac{\partial}{\partial w}\left[(1-w^2)\frac{\partial U}{\partial w}\right]=\frac{\partial}{\partial \mu}\left[(1-\mu^2)\frac{\partial U}{\partial \mu}\right]+\frac{1}{1-\mu^2}\frac{\partial^2 U}{\partial \phi^2}$$

which occurs in Poisson's work on the theory of sound. The integrals derived in the present paper are closely associated with this differential equation and give some of the ground work for a complete discussion of it. They have been evolved gradually from an initial clue suggested by some work of Beltrami on symmetrical potential functions, one author making one step and the other the next and so on.

Outline of results. The comparison of two types of integrals, due to Beltrami and Laplace respectively, for symmetrical potential functions suggests the existence of the relations stated in equations (4) and (5). These equations may then be verified directly. It is also possible to establish them in a modified form by complex integration.

These identities are suitable for dealing with Legendre functions, and when applied to an integral which follows from Hobson's addition theorem for $Q_n(z)$, they lead to the integrals given by equations (10) and (11). By analytic continuation of the parameters in these integrals still other integrals are derived. Equation (22) was first obtained by applying the recurrence relations for Legendre functions to the four integrals (14), (15), (18) and (19). However, this work is omitted, as here it is more convenient to regard (22) as a special case of equation (35).

Up to and including equation (22) the degree n and order m of every Legendre function are integers. A study of equation (22) with the idea of extending it to non-integer values of n and m led to theorems concerning contour-integrals having Legendre functions in the integrand. An application of one of these theorems leads to the desired extension given by equation (35). The theorem also suggests the result given by equation (36).

The paper ends with the derivation of an expansion for which (22) is

---

* Received December 6, 1937.

used to determine the coefficients. In the accompanying paper the existence of this expansion is suggested by solutions of the wave equation.

Potential functions having radial symmetry about an axis. If x, y, z are rectangular coördinates and $w = \sqrt{x^2 + y^2}$ then a potential function symmetrical about the z axis is obtained by integrating the elementary potential $1/R$[1]

$$V = \int_{-a}^{a} \frac{F(t)\,dt}{R} \tag{1}$$

where the constant a and the function $F(t)$ are arbitrary and

$$R^2 = w^2 + (z + it)^2.$$

Another symmetrical potential involving R may be obtained from Laplace's expression for a potential which assumes the value $F(z)$ along the Z axis,

$$V = \frac{1}{\pi} \int_0^{\pi} f(z + iw \cos k)\,dk.$$

Changing the variable leads to

$$V = \frac{1}{\pi} \int_{iz-w}^{iz+w} \frac{f(-it)\,dt}{R}. \tag{2}$$

The similarity of the integrands suggests the existence of a relationship between the functions $F(t)$ and $f(-it)$. Consideration of the potential of a circular ring of radius a indicated that this relation is

$$f(z) = \int_{-a}^{+a} \frac{F(t)\,dt}{z + it}. \tag{3}$$

That is, when (3) is satisfied, we expect the equality

$$\int_{-a}^{a} dt \frac{F(t)}{R} = \frac{1}{\pi} \int_{iz-w}^{iz+w} \frac{f(-it)}{R}\,dt. \tag{4}$$

This may be verified by substituting the expression for $f(z)$ given by (3) in the right member and integrating after inverting the order of integration.

If instead of expressions (1) and (2) we start with

---

[1] This expression is closely associated with Beltrami's integrals for a symmetrical potential.

Rendiconti Lombardo (2), vol. 11 (1878), pp. 668-680 = *Opere,* vol. 3, pp. 115-128.
Bologna Mem. (4), vol. 2 (1880), pp. 461-505 = *Opere,* vol. 3, pp. 349-382.
Bologna Mem. (4), vol. 4 (1882), pp. 211-216 = *Opere,* vol. 4, pp. 45-76.

$$W = \int_{-a}^{+a} F(t)\,(z + it)\,dt/R$$

$$W = \frac{iw}{\pi}\int_0^{\pi} f(z + iw\cos k)\cos k\,dk,$$

the first of which is related to Beltrami's integrals, we obtain in place of (4)

$$(5)\quad \int_{-a}^{a} \frac{F(t)\,(z+it)\,dt}{R} - \int_{-a}^{a} F(t)\,dt = \frac{1}{\pi}\int_{iz-w}^{iz+w} \frac{(z+it)f(-it)\,dt}{R}$$

which may be verified in the same way as (4).

It is also possible to establish these identities, in a modified form, by using Cauchy's theorem. The result may be stated as follows:

Let $g(s)$ be continuous on the path L in the s-plane and let

$$(6)\qquad h(t) = \frac{1}{2\pi i}\int_L \frac{g(s)\,ds}{s-t}$$

then, if $S(t) = (t-t_1)(t-t_2)$, we have the relation

$$(7)\qquad \frac{1}{\pi}\int_{t_1}^{t_2} [-S(t)]^{-\frac{1}{2}} h(t)\,dt = \frac{1}{2\pi i}\int_D [S(s)]^{-\frac{1}{2}} g(s)\,ds$$

provided D does not cross the line joining t_1 and t_2. The arguments are chosen so that

$$\arg(t-t_1) = \arg(t_2-t) = \theta,$$

where θ is the angle which the vector from t_1 to t_2 makes with the real axis, and

$$\theta \leq \arg(s-t_1) < 2\pi + \theta, \qquad 0 \leq \arg(s-t_2) < 2\pi.$$

Under these conditions there is also an analogue of (5), namely

$$(8)\qquad \frac{1}{\pi}\int_{t_1}^{t_2} [-S(t)]^{-\frac{1}{2}} h(t)\,t\,dt$$
$$= -\frac{1}{2\pi i}\int_D g(s)\,ds + \frac{1}{2\pi i}\int_D [S(s)]^{-\frac{1}{2}} g(s)\,s\,ds.$$

The first step in the proof of (7) is to replace $\frac{1}{\pi}\int_{t_1}^{t_2}(\quad)\,dt$ by $\frac{1}{2\pi i}\int_C(\quad)\,dt$ where C is a contour enclosing the points t_1 and t_2 which does not cross D. Use of (6) and a change of order of integration, which may be justified, leads to the integral

$$\frac{1}{2\pi i}\int_C \frac{[S(t)]^{-\frac{1}{2}}\,dt}{s-t}$$

which may be evaluated by expanding C until it consists of an infinite circle plus a loop about $t = s$. Only the latter contributes to the value of the integral, which is seen to be $[S(s)]^{-\frac{1}{2}}$. The result stated is then obtained by letting C, on the other side of the equation, shrink to a dumbell shaped contour with very small loops around t_1 and t_2.

Equation (8) is proved in an analogous manner.

Application of the identities to integrals involving Legendre functions. The fact that Legendre's functions of the first and second kinds are related, for integer values of n, by

$$Q_n(t) = \frac{1}{2}\int_{-1}^{1} \frac{P_n(s)\,ds}{t-s}$$

enables us to apply the identities to these functions. For upon setting $h(t) = Q_n(t)$, $g(s) = -\pi i P_n(s)$ and taking D to be the line joining -1 and $+1$ we see that equation (6) becomes the one just given. To obtain an integral of the type used in the identities we note from Hobson's addition theorem for $Q_n(z)$ [2] it follows that

$$Q_n(b)P_n(c) = \frac{1}{\pi}\int_0^{\pi} Q_n(bc - (b^2-1)^{\frac{1}{2}}(c^2-1)^{\frac{1}{2}}\cos a)\,da$$

where n is an integer and $b > c > 1$. Changing the variable of integration transforms the integral into

$$Q_n(b)P_n(c) = \frac{1}{\pi}\int_{t_1}^{t_2} \frac{Q_n(t)\,dt}{\sqrt{(t_2-t)(t-t_1)}} \tag{9}$$

where

$$t_2 = bc + \sqrt{(b^2-1)(c^2-1)}, \qquad t_1 = bc - \sqrt{(b^2-1)(c^2-1)}.$$

Since $t_2 > t_1 > 1$, the path D joining -1 and $+1$ does not cross the line joining t_2 and t_1. Also $P_n(s)$ is continuous on D and we may apply equation (7) to transform the integral in equation (9). Thus noting that $\theta = 0$ and $\arg\,(s - t_1) = \arg\,(s - t_2) = \pi$ we obtain the known result

$$\begin{aligned} 2Q_n(b)P_n(c) &= \int_{-1}^{+1} \frac{P_n(s)\,ds}{\sqrt{(t_2-s)(t_1-s)}} \\ &= \int_{-1}^{+1} \frac{P_n(s)\,ds}{\sqrt{b^2+c^2+s^2-2bcs-1}}. \end{aligned} \tag{10}$$

---

[2] E. W. Hobson, *The Theory of Spherical and Ellipsoidal Harmonics,* Cambridge (1931), p. 378. In the following work this text will be denoted by Hobson (1).

In the same way we obtain from (8) and

$$Q_n^1(b)P_n^{-1}(c) = \frac{1}{\pi}\int_0^\pi Q_n(bc - \sqrt{(b^2-1)(c^2-1)}\cos a)\cos a\,da$$

the equation

$$(11)\quad 2Q_n^1(b)P_n^{-1}(c) = \int_{-1}^{+1} \frac{(s-bc)P_n(s)\,ds}{\sqrt{(b^2-1)(c^2-1)}\sqrt{b^2+c^2+s^2-2bcs-1}}$$

which holds when n is a positive integer.

Analytic continuation applied to the integrals in equations (10) and (11). When the integrals in equations (10) and (11) are viewed as analytic functions of the parameters b and c, some interesting integrals may be obtained by continuing b and c from their original values, which were greater than unity, to values lying between -1 and $+1$. For the time being we shall denote these new values by b_1 and c_1 respectively and assume that $b_1 > c_1$.

In the expressions for t_1 and t_2 we first let c decrease to c_1 along a path which lies on the real axis except for an indentation upward at unity. Then the corresponding values of t_1 and t_2 are

$$t_2 = bc_1 + i\sqrt{(b^2-1)(1-c_1^2)}, \qquad t_1 = bc_1 - i\sqrt{(b^2-1)(1-c_1^2)}$$

When we let b decrease to b_1 along a similar path we obtain

$$t_2 = b_1c_1 - \sqrt{(1-b_1^2)(1-c_1^2)}, \qquad t_1 = b_1c_1 + \sqrt{(1-b_1^2)(1-c_1^2)}$$

where now both t_1 and t_2 lie between -1 and $+1$ and $t_1 > t_2$. Incidentally, the reason for keeping $b_1 > c_1$ becomes apparent upon letting $b_1 = c_1$ for then $t_2 = -1$ and $t_1 = +1$.

We must take care that in this process neither t_2 nor t_1 crosses the path of integration D of s in (10) and (11). A more detailed examination of the paths followed by t_2 and t_1 shows that t_2 must lie above and t_1 below D. For this reason D is indented downward at $s = t_2$ and upward at $s = t_1$.

Since the path followed by b in decreasing to b_1 was indented so as to pass above the point $+1$, b_1 will lie above the cut in the z plane joining -1 to $+1$ associated with $Q_n(z)$. Thus, following the usual custom, we write $Q_n(b_1 + i0)$.

Splitting the integral (10) up into its real and imaginary parts we obtain, upon dropping the subscript one,

$$(12)\quad 2Q_n(b+i0)P_n(c) = \left\{\int_{-1}^{t_2} + \int_{t_1}^{1}\right\} \frac{ds\,P_n(S)}{\sqrt{b^2+c^2+s^2-2bcs-1}} - i\int_{t_2}^{t_1} \frac{ds\,P_n(s)}{\sqrt{1+2bcs-b^2-c^2-s^2}}$$

where now $-1 < c < b < 1$. The function $Q_n(b + i0)$ may be split into its real and imaginary parts by utilizing the definition of $Q_n(z)$ when $-1 < z < 1$;

$$Q_n(z) = Q_n(z + i0) + \frac{\pi i}{2} P_n(z).$$

Equating imaginary parts gives, after a transformation inverse to that used in obtaining equation (9), the result

$$P_n(b) P_n(c) = \frac{1}{\pi} \int_0^\pi P_n[bc - (1 - b^2)^{\frac{1}{2}} (1 - c^2)^{\frac{1}{2}} \cos a] \, da.$$

This equation may be readily obtained from the addition theorem for Legendre polynomials.

Setting $b = \tanh u$, $c = \tanh v$ and equating imaginary parts gives

$$\begin{aligned} (13) \qquad & 2 Q_n(\tanh u) P_n(\tanh v) \\ & = \int_0^{u+v} P_n\left(\tanh u \tanh v - \frac{ch\, t}{ch\, u\, ch\, v}\right) dt \\ & + \int_0^{u-v} P_n\left(\tanh u \tanh v + \frac{ch\, t}{ch\, u\, ch\, v}\right) dt \end{aligned}$$

where it is assumed that $u > v > 0$ in the derivation. Since both sides are analytic functions of v in the neighborhood of the real axis this restriction may be removed by analytic continuation.

The following are special cases of equation (13)

$$(14) \qquad Q_{2n+1}(0) P_{2n+1}(\tanh v) = - \int_0^v P_{2n+1}\left(\frac{ch\, t}{ch\, v}\right) dt$$

$$(15) \qquad Q_{2n}(\tanh u) P_{2n}(0) = \int_0^u P_{2n}\left(\frac{ch\, t}{ch\, u}\right) dt$$

$$Q_n(\tanh u) P_n(\tanh u) = \int_0^u P_n\left(1 - 2 \frac{ch^2\, s}{ch^2\, u}\right) ds.$$

In the same way equation (11) leads to the analogue of (13)

$$\begin{aligned} (16) \qquad & 2 \text{ Real Part of } Q_n^{1}(\tanh u + i0) P_n^{-1}(\tanh v + i0) \\ & = \int_0^{u-v} P_n\left(\tanh u \tanh v - \frac{ch\, t}{ch\, u\, ch\, v}\right) ch\, t\, dt \\ & - \int_0^{u-v} P_n\left(\tanh u \tanh v + \frac{ch\, t}{ch\, u\, ch\, v}\right) ch\, t\, dt. \end{aligned}$$

By using the following results[3] which hold for general values of m and n

---

[3] The first two are given by E. W. Hobson, *Philosophical Transactions*, vol. 187 (1896), pp. 443-531, equations (19) and (29). The third one is given Hobson (1), p. 229.

$$(17)\quad P_n^{-m}(x+iy)=\frac{\Pi(n-m)}{\Pi(n+m)}\left\{ P_n^{m}(x+iy)-\frac{2}{\pi}e^{-m\pi i}\sin m\pi\, Q_n^{m}(x+iy)\right\}$$

$$P_n^{m}(x+i0)=e^{-m\pi i/2}P_n^{m}(x)$$

$$Q_n^{m}(x+i0)=e^{3m\pi i/2}\left\{ Q_n^{m}(x)-\frac{i\pi}{2}P_n^{m}(x)\right\}$$

where x and y are real numbers, the left-hand side of (16) may be written as

$$-\frac{2}{n(n+1)}Q_n^{1}(\tanh u)P_n^{1}(\tanh v).$$

Special cases of (16) are

$$(18)\quad -\frac{1}{2n(2n+1)}Q_{2n}^{1}(0)P_{2n}^{1}(\tanh v)=\int_0^v P_{2n}\left(\frac{ch\, t}{ch\, v}\right)ch\, t\, dt$$

$$(19)\quad \frac{1}{(2n+1)(2n+2)}Q^{1}_{2n+1}(\tanh u)P^{1}_{2n+1}(0)=\int_0^u P_{2n+1}\left(\frac{ch\, t}{ch\, u}\right)ch\, t\, dt$$

$$(20)\quad -\frac{1}{n(n+1)}Q_n^{1}(\tanh u)P_n^{1}(\tanh u)=\int_0^u P_n\left(1-\frac{2ch^2 s}{ch^2 u}\right)ch\, 2s\, ds.$$

By applying the recurrence relation for the Legendre functions to equations (14), (15), (18), (19) and using the expressions [4]

$$(21)\quad \begin{aligned} P_n^{m}(0)&=2^{m}\cos\frac{n+m}{2}\pi\,\frac{\Pi\left(\frac{n+m-1}{2}\right)}{\Pi\left(\frac{n-m}{2}\right)\Pi(-\frac{1}{2})}\\ Q_n^{m}(0)&=-2^{m-1}\sin\frac{n+m}{2}\pi\,\frac{\Pi\left(\frac{n+m-1}{2}\right)\Pi(-\frac{1}{2})}{\Pi\left(\frac{n-m}{2}\right)}\end{aligned}$$

the following relation may be established

$$(22)\quad \int_0^v P_n\left(\frac{ch\, u}{ch\, v}\right)ch\, mu\, du=\begin{cases}-Q_n^{n}(\tanh v)/Q_n^{m+1}(0), & m+n \text{ even}\\ -P_n^{-n}(\tanh v)/P_n^{m+1}(0), & m+n \text{ odd.}\end{cases}$$

In this equation m and n are assumed to be positive integers.

Further integrals involving Legendre functions. Equation (22) suggests the problem of determining the value of the integral when m and n are not integers. Since none of the foregoing work may be applied when m and n are not integers we must search for a different method of attack. A study of (22) suggests the two theorems

---

[4] Hobson (1), p. 232.

1. If $\phi(x)$ be a Legendre function of order n, the integral

$$I = \int_P \frac{\phi(au)\exp(m\, ch^{-1}u)\, du}{\sqrt{u^2-1}} \tag{23}$$

may be expressed in terms of the associated Legendre functions:

$$I = AP_n^m(b) + BQ_n^m(b) \tag{24}$$

where $b = \sqrt{1-a^2}$ and A and B are independent of a, provided the expression

$$F(u) = [a\sqrt{u^2-1}\,\phi'(au) - m\phi(au)]\exp(m\, ch^{-1}u) \tag{25}$$

has the same value at the initial and final points, both assumed independent of a, of the path P. The prime denotes differentiation: $\phi'(t) = \dfrac{d\phi(t)}{dt}$.

2. The same is true of the integral

$$J = \int_P \frac{\phi(t)\exp\left(m\, ch^{-1}\dfrac{t}{a}\right) dt}{\sqrt{t^2-a^2}} \tag{26}$$

provided the expression

$$G(t) = (t^2-1)\left[\frac{\phi'(t)}{z} + \frac{t\phi(t)}{z^3} - \frac{m\phi(t)}{z^2}\right]\exp\left(m\, ch^{-1}\frac{t}{a}\right) \tag{27}$$

where $z = \sqrt{t^2-a^2}$, assumes the same value at the initial and final points, which are here also taken to be independent of a, of the path P.

Although these results were found more or less by experiment, they may be verified by straightforward differentiation. Thus it may be shown that when I and J are placed in the differential equation for the associated Legendre functions we obtain

$$\frac{d}{db}(1-b^2)\frac{dI}{db} + \left\{n(n+1) - \frac{m^2}{1-b^2}\right\} I = \frac{1}{a^2}\int_P du \frac{d}{du} F(u)$$
$$\frac{d}{db}(1-b^2)\frac{dJ}{db} + \left\{n(n+1) - \frac{m^2}{1-b^2}\right\} J = \int_P dt \frac{d}{dt} G(t)$$

and if $F(u)$ and $G(t)$ assume the same values at the initial and final points of their respective paths the integrals on the right vanish.

Since m occurs in the differential equation as m^2 we may change the sign of m in I and J. Addition and subtraction will then give theorems concerning integrals having hyperbolic cosines and sines in place of the exponential functions in I and J. In particular we obtain the result that

$$K = \int_P \phi(t)\, ch\left(m\, ch^{-1}\frac{t}{a}\right)\frac{dt}{\sqrt{t^2 - a^2}} \tag{28}$$

may be expressed in the form (24) provided

$$H(t) = (t^2 - 1)\left[\left(\frac{\phi'}{z} + \frac{t\phi}{z^3}\right) ch\left(m\, ch^{-1}\frac{t}{a}\right) - \frac{m\phi}{z^2}\, sh\left(m\, ch^{-1}\frac{t}{a}\right)\right] \tag{29}$$

assumes the same value at the initial and final points of P.

In order to generalize equation (22) we consider the expression (28) where $\phi(t) = P_n(t)$ and P is a path which starts at $t = 1$, passes around $t = a$ in the positive direction, and returns to the point $t = 1$. For the sake of convenience we assume that a is real, $0 < a < 1$, and $\arg(t - a) = \arg(t + a) = 0$ at the starting point. The expression (29) vanishes at both ends of P, because of the factor $t^2 - 1$, and hence K may be expressed in the form (24). Instead of K it is more suitable to deal with L defined by

$$L = \int_a^1 P_n(t)\, ch\left(m\, ch^{-1}\frac{t}{a}\right)\frac{dt}{\sqrt{t^2 - a^2}} \tag{30}$$

where the square root is supposed to be positive.

By shrinking P to a small loop about $t = a$ plus two straight lines joining the circle to $t = 1$ we see that

$$L = -\frac{1}{2}K$$

and thus L may also be expressed as

$$L = AP_n^m(\sqrt{1 - a^2}) + BQ_n^m(\sqrt{1 - a^2}). \tag{31}$$

As $a \to 1$, $L \to 0$ since $ch^{-1}\left(\frac{1}{a}\right) \to 0$ and $P_n(1) = 1$. Therefore

$$0 = AP_n^m(0) + BQ_n^m(0) \tag{32}$$

is one relation between A and B. To obtain a second relation we multiply the numerator and denominator of (30) by t and use

$$\frac{t\, dt}{\sqrt{t^2 - a^2}} = d(\sqrt{t^2 - a^2})$$

to integrate by parts. This leads to

$$\frac{dL}{da} \to -\frac{1}{\sqrt{1 - a^2}} \text{ as } a \to 1.$$

When this is placed in (31) after that equation has been differentiated, cancellation of the common factor $\frac{1}{\sqrt{1-a^2}}$ gives, upon letting $a \to 1$,

$$(33) \qquad 1 = \left[A \frac{d}{dx} P_n{}^m(x) + B \frac{d}{dx} Q_n{}^m(x) \right]_{x=0}.$$

Equations (32) and (33) may now be solved for A and B. If C denotes the constant in

$$P_n{}^m(x) \frac{d}{dx} Q_n{}^m(x) - Q_n{}^m(x) \frac{d}{dx} P_n{}^m(x) = \frac{C}{1-x^2}.$$

The value of L is

$$(34) \qquad L = \frac{1}{C} \left[P_n{}^m(0) Q_n{}^m(\sqrt{1-a^2}) - Q_n{}^m(0) P_n{}^m(\sqrt{1-a^2}) \right].$$

The values of $P_n{}^m(0)$ and $Q_n{}^m(0)$ are given by (21), while[5]

$$C = 2^{2m} \frac{\Pi\left(\frac{n+m-1}{2}\right) \Pi\left(\frac{n+m}{2}\right)}{\Pi\left(\frac{n-m-1}{2}\right) \Pi\left(\frac{n-m}{2}\right)}.$$

When Hobson's definitions of the associated Legendre functions for $-1 < x < 1$ are used, substitution of the values for C, $P_n{}^m(0)$, $Q_n{}^m(0)$ in (34) and the change of variable $t = a\, ch\, u$, $ch\, v = 1/a$ in (30) gives for the final result

$$(35) \qquad \int_0^v P_n\left(\frac{ch\, u}{ch\, v}\right) ch\, m\, u\, du$$
$$= \frac{\sqrt{\pi}\, \Pi\left(\frac{n-m-1}{2}\right)}{2^m\, \Pi\left(\frac{n+m}{2}\right)} \left[\frac{\cos(n+m)\frac{\pi}{2}}{\pi} Q_n{}^m(\tanh v) + \frac{\sin(n+m)\frac{\pi}{2}}{2} P_n{}^m(\tanh v) \right]$$

which is a generalization of (22).

In a somewhat similar manner it may be shown that

$$(36) \qquad \int_0^\infty Q_n(a\, ch\, u)\, ch\, m\, u\, du = Q_n{}^{-m}(0+i0)\, Q_n{}^m(-i\sqrt{a^2-1})$$
$$= \frac{\sqrt{\pi}\, \Pi\left(\frac{n-m-1}{2}\right)}{2^{m+1}\, \Pi\left(\frac{n+m}{2}\right)} e^{-i\pi[m+(n+1)/2]} Q_n{}^m(-i\sqrt{a^2-1})$$

---

[5] Hobson (1), p. 232.

where $R(n+1-m)>0$, $R(n+m+1)>0$, which imply $R(n)>-1$; and $\arg(\sqrt{a^2-1})=0$ when $\arg a=0$ and $|a|>1$. Equation (36) is suggested by (28) when $\phi(t)$ is set equal to $Q_n(t)$ and P is taken to be a path starting and ending at $+\infty$ which comes in and encloses $t=a$. It may be readily verified when $|a|>1$ by replacing $Q_n(a\,ch\,u)$ by its series expansion in powers of $1/a\,ch\,u$ and integrating termwise.

Expansions suggested by equation (22). Since a function which satisfies Darboux's conditions[6] may be expanded in the Legendre series[7]

$$f(z)=\sum_{n=0}^{\infty}a_nP_n(z)\,(n+\tfrac{1}{2}) \tag{37}$$

where

$$a_n=\int_{-1}^{+1}f(t)P_n(t)\,dt, \tag{38}$$

and since (22) is of the form (38), it is natural to seek the corresponding function $f(z)$. By setting $t=chu/chv$, $x=\operatorname{sech}v$ we see that the integral in (22) is changed into the integral occurring in (30) and hence we are led to set

$$\begin{aligned} f(t)&=ch\left(m\,ch^{-1}\frac{t}{a}\right)/\sqrt{t^2-a^2}=T_m(t/a)/\sqrt{t^2-a^2}, && a<t<1\\ &=0 && ,\ -1<t<a \end{aligned}$$

where $T_m(z)$ is Tchebycheff's polynomial which is such that when $z=ch\,u$, $T_m(z)=ch\,mu$. When m is even, say equal to $2k$, the expansion written in full is

$$\left.\begin{aligned} &\operatorname{sech}v<t<1, \frac{T_{2k}(t\,ch\,v)}{\sqrt{t^2-\operatorname{sech}^2v}}\\ &-1<t<\operatorname{sech}v,\qquad 0 \end{aligned}\right\}=\begin{aligned} &-\sum_{n=0}^{\infty}P_{2n}(t)\,(2n+\tfrac{1}{2})\,Q_{2n}^{2k}(\tanh v)/Q_{2n}^{2k+1}(0)\\ &-\sum_{n=0}^{\infty}P_{2n+1}(t)\,(2n+\tfrac{3}{2})\,P_{2n+1}^{2k}(\tanh v)/P_{2n+1}^{2k+1}(0). \end{aligned} \tag{39}$$

The first series represents an even function of t whereas the second one represents an odd function. When t is negative the sum of the two functions is zero. Hence when t is positive the two functions are equal. Separating the even and odd portions of (39) gives

$$\begin{aligned} &\sum_{n=0}^{\infty}(4n+1)P_{2n}(t)\,Q_{2n}^{2k}(\tanh v)/Q_{2n}^{2k+1}(0)=-\frac{T_{2k}(t\,ch\,v)}{\sqrt{t^2-\operatorname{sech}^2v}}\text{ or }0\\ &\sum_{n=0}^{\infty}(4n+3)P_{2n+1}(t)\,P_{2n+1}^{2k}(\tanh v)/P_{2n+1}^{2k+1}(0)=-\frac{T_{2k}(t\,ch\,v)}{t\sqrt{1-t^{-2}\operatorname{sech}^2v}}\text{ or }0. \end{aligned} \tag{40}$$

---

[6] G. Darboux, *Journal de Mathématique* (3), vol. 4 (1878), pp. 5-56, 377-416.

[7] E. W. Hobson, *Proceedings of the London Mathematical Society* (2), vol. 7 (1909), pp. 24-39.

The first or second value on the right is to be taken accordingly as $t^2 - \operatorname{sech}^2 v$ is positive or negative. The positive value of the square root is to be used. By considering m to be odd the expansions

$$\begin{aligned}(41)\quad &\sum_{n=0}^{\infty}(4n+1)P_{2n}(t)P_{2n}^{2k+1}(\tanh v)/P_{2n}^{2k+2}(0) = -\frac{T_{2k+1}(t\,ch\,v)}{\sqrt{t^2-\operatorname{sech}^2 v}} \text{ or } 0\\ &\sum_{n=0}^{\infty}(4n+3)P_{2n+1}(t)Q_{2n+1}^{2k+1}(\tanh v)/Q_{2n+1}^{2k+2}(0) = -\frac{T_{2k+1}(t\,ch\,v)}{t\sqrt{1-t^{-2}\operatorname{sech}^2 v}} \text{ or } 0\end{aligned}$$

are obtained.

CALIFORNIA INSTITUTE OF TECHNOLOGY.
AND
BELL TELEPHONE LABORATORIES.

A NOTE ON AN EXTENSION OF BERNSTEIN'S THEOREM.*

By W. H. McEwen.

In a recent paper[1] the author obtained an extension of Bernstein's theorem for arbitrary sums of characteristic solutions of a general n-th order linear differential system

$$(1)\qquad \begin{aligned} &L(u) \equiv u^{(n)} + P_2(x)u^{(n-2)} + \cdots + P_n(x)u + \lambda u = 0, \\ &W_j(u) = 0 \qquad\qquad\qquad\qquad (j = 1, 2, \cdots, n), \end{aligned}$$

in which the coefficients $P_j(x)$ are continuous with continuous derivatives of all orders on (a, b), the boundary conditions are normalized and regular on (a, b), and the complex characteristic values λ_k (arranged in order of increasing moduli) give rise to poles of the Green's function which are simple when k is large.

The sums in question have the form $S_N(x) = \sum_{i=1}^{N} a_i u_i(x)$, in which the a's are arbitrary and the u's are the characteristic solutions of (1) corresponding respectively to the first N characteristic values $\lambda_1, \lambda_2, \cdots, \lambda_N$. Assuming that $| S_N(x) | \leqq L$ on (a, b) it was found that $| S'_N(x) | \leqq QNL$ uniformly on any interior interval $a + \delta \leqq x \leqq b - \delta$, Q being a constant independent of N, and an example was cited to show that this is the best result that can be obtained in general. In particular cases, however (as for example in the Fourier case or the Sturm-Liouville case), the limit QNL can be applied to the whole interval (a, b). In the present paper we propose to investigate further the circumstances under which this extension to the whole interval can be made. The discussion will be based on results found in a paper by Stone[2] (referred to hereafter as (S)), and the author's paper[3] (referred to as (M)).

Let us define, over the range $a \leqq a' \leqq b' \leqq b$, the function $Q(a', b')$, $0 \leqq Q(a', b') \leqq +\infty$, as the "best" constant such that

$$| S'_N(x) | \leqq Q(a', b')NL$$

---

* Received September 28, 1937.

[1] W. H. McEwen, "An extension of Bernstein's theorem associated with general boundary value problems," *American Journal of Mathematics*, vol. 59 (1937), pp. 295-305.

[2] M. H. Stone, "A comparison of the series of Fourier and Birkhoff," *Transactions of the American Mathematical Society*, vol. 28 (1926), pp. 695-761.

[3] *Loc. cit.*

for $a' \leqq x \leqq b'$, where L is the maximum value of $|S_N|$ on $a \leqq x \leqq b$. It is readily seen that the function Q has the following properties:

(i) if $a \leqq a'' \leqq a' \leqq b' \leqq b'' \leqq b$, then $Q(a', b') \leqq Q(a'', b'')$;
(ii) $Q(a', b') \to Q(a'', b'')$ when $a' \to a''$ and $b' \to b''$;
(iii) if $a \leqq a' \leqq b' \leqq c' \leqq b$, then $Q(a', c') = \max[Q(a', b'), Q(b', c')]$;
(iv) if $a < a' \leqq b' < b$, then $Q(a', b') < +\infty$.

The first three are implied by the definition of Q as the "best" constant, whereas the fourth was established in the author's earlier paper. From these properties it is evident that any further interest in the function Q centers in its behaviour as $a' \to a$ or $b' \to b$.

We can assume, without loss in generality, that the interval of x is $0 \leqq x \leqq 1$, and the maximum value of $|S_N(x)|$ on $(0,1)$ is 1. The boundary conditions of (1), being normalized, can be written

$$W_j(u) \equiv \alpha_j u^{(k_j)}(0) + \beta_j u^{(k_j)}(1) + \sum_{i=0}^{k_j-1} (\alpha_{ij} u^{(i)}(0) + \beta_{ij} u^{(i)}(1)) = 0, \qquad (j = 1, 2, \cdots, n),$$

in which the k_j's are positive integers such that $n - 1 \geqq k_1 \geqq k_2 \geqq \cdots \geqq k_n$ and no three k_j's are the same. Along with (1) let us consider a second system of the same type and of the same order n:

$$\begin{aligned} &\bar{L}(u) \equiv u^{(n)} + \bar{P}_2(x) u^{(n-2)} + \cdots + \bar{P}_n(x) u + \lambda u = 0, \\ (2) \qquad &\bar{W}_j(u) \equiv \bar{\alpha}_j u^{(\bar{k}_j)}(0) + \bar{\beta}_j u^{(\bar{k}_j)}(1) + \sum_{i=0}^{k_j-1} (\bar{\alpha}_{ij} u^{(i)}(0) + \bar{\beta}_{ij} u^{(i)}(1)) = 0, \\ &\qquad (j = 1, 2, \cdots, n), \end{aligned}$$

and define

Hypothesis A. Systems (1) and (2) are so related that $\alpha_j = \bar{\alpha}_j$, $\beta_j = \bar{\beta}_j$ $k_j = \bar{k}_j$, $(j = 1, 2, \cdots, n)$.

Let $G(x, y; \lambda)$ and $\bar{G}(x, y; \lambda)$ be the Green's functions respectively of systems (1) and (2), and let $\lambda = \rho^n$. The sum $S_N(x)$ may then be written as a contour integral

$$S_N(x) = \frac{1}{2\pi i} \int_0^1 S_N(y) \int_\Gamma n\rho^{n-1} G(x, y; \rho^n) d\rho dy,$$

where Γ is an arc of the circle $|\rho| = R$ in the complex ρ-plane and is defined by two adjacent sectors of the set of $2n$ equal sectors:

$$l\pi/n \leqq \arg \rho \leqq (l+1)\pi/n \qquad (l = 0, 1, 2, \cdots, 2n-1).$$

The radius of Γ, $R \sim \pi N$, and is adjusted so that the arc remains uniformly away from the poles of both G and $\bar{G}$ when R (or N) is large.[4] The function

---

[4] See (M), p. 297.

$$\bar{S}_N(x) = \frac{1}{2\pi i}\int_0^1 S_N(y) \int_\Gamma n\rho^{n-1}\bar{G}(x, y; \rho^n)\, d\rho\, dy$$

will then represent the partial sum of order $\bar{N}$ of the Birkhoff expansion of the function $S_N(x)$ associated with system (2). Moreover the order $\bar{N}$ will satisfy $| N - \bar{N} | \leqq K$ so that $0(\bar{N}) = 0(N)$.

We now observe two lemmas:

LEMMA 1. *Under hypothesis A,*

$$\int_\Gamma n\rho^{n-1}(G - \bar{G})\, d\rho = 0(1)$$

uniformly on $0 \leqq x, y \leqq 1$.

LEMMA 2. *Under hypothesis A,*

$$\int_\Gamma n\rho^{n-1}\left(\frac{\partial G}{\partial x} - \frac{\partial \bar{G}}{\partial x}\right) d\rho = 0(N)$$

uniformly on $0 \leqq x, y \leqq 1$.

These lemmas are extensions respectively of lemmas 2 and 3 of (M), and are obtained from the latter by imposing the additional restrictions contained in hypothesis A. A proof of Lemma 1 for the case of a system of odd order is given in (S, Theorem XV, pp. 729-730). An analogous argument will suffice for the case of even order. The nature of the argument involved will be brought out in our outline of the proof of Lemma 2, for the case of odd order, $n = 2\mu - 1$, which follows:

From (M, p. 302) we obtain the formula

$$(3) \qquad n\rho^{n-1}\left(\frac{\partial G}{\partial x} - \frac{\partial \bar{G}}{\partial x}\right) \equiv \left\{ \begin{array}{l} (F^0{}_{11} - \bar{F}^0{}_{11}) - \sum_{j=1}^{\mu} e^{\rho\omega_j(x-y)} \dfrac{(m_j - \bar{m}_j)}{\rho}\,; \\ (F^1{}_{11} - \bar{F}^1{}_{11}) + \sum_{j=\mu+1}^{n} e^{\rho\omega_j(x-1)} \dfrac{(m_j - \bar{m}_j)}{\rho} \end{array} \right\}$$
$$+ \sum_{j=1}^{\mu} (\rho\omega_j) e^{\rho\omega_j x}(E_j - \bar{E}_j) + \sum_{j=\mu+1}^{n} (\rho\omega_j) e^{\rho\omega_j(x-1)}(E_j - \bar{E}_j),$$

which holds for values of ρ on the arc γ^1 (γ^1 is one of the halves of γ, which in turn is one of the halves of Γ. See M, p. 299). The functions $(F^0{}_{11} - \bar{F}^0{}_{11})$, $(F^1{}_{11} - \bar{F}^1{}_{11})$, m_j, $\bar{m}_j$, E_j, $\bar{E}_j$ and all the exponentials are uniformly bounded on $0 \leqq x, y \leqq 1$ as $R \to \infty$ (see M, p. 302). Hence the expression $\{ \ \} = 0(1)$. On the other hand the summations in the last line of (3) are in general $0(R) = 0(N)$. However, under hypothesis A these also become $0(1)$, for in that case we find that $E_j - \bar{E}_j = 0(1)/\rho$.

To prove this last statement we note that the summations in question arise from the evaluation of the expression

$$(4)\qquad \frac{\Delta_1^{(1)}}{[\theta_1]e^{\rho\omega_\mu}+[\theta_0]}-\frac{\bar{\Delta}_1^{(1)}}{[\bar{\theta}_1]e^{\rho\omega_\mu}+[\bar{\theta}_0]},$$

in which the denominators are uniformly bounded away from zero as $R\rightarrow\infty$, and

$$\Delta_1^{(1)}\equiv\begin{vmatrix} (\rho\omega_1)e^{\rho\omega_1 x}[1] & \cdot & \cdot & \cdot & (\rho\omega_n)e^{\rho\omega_n(x-1)}[1] & 0 \\ [\alpha_1\omega_1^{k_1}] & \cdot & \cdot & \cdot & [\beta_1\omega_n^{k_1}] & D_1 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ [\alpha_n\omega_1^{k_n}] & \cdot & \cdot & \cdot & [\beta_n\omega_n^{k_n}] & D_n \end{vmatrix}^{5},$$

$$D_i\equiv\sum_{j=1}^{\mu}e^{\rho\omega_j(1-y)}[\beta_i\omega_j^{k_i+1}]-\sum_{j=\mu+1}^{n}e^{-\rho\omega_j y}[\alpha_i\omega_j^{k_i+1}],$$

and $\bar{\Delta}_1^{(1)}$ is a similar form involving $\bar{\alpha}_j$, $\bar{\beta}_j$, $\bar{k}_j$. On expanding each determinant according to the elements of its first row and collecting terms in (4) we obtain the last line of (3). Hence we may write

$$E_j=\frac{M_j[1]}{[\theta_1]e^{\rho\omega_\mu}+[\theta_0]},\qquad \bar{E}_j=\frac{\bar{M}_j[1]}{[\bar{\theta}_1]e^{\rho\omega_\mu}+[\bar{\theta}_0]},$$

where M_j, $\bar{M}_j$ are the cofactors of the j-th elements in the first rows of $\Delta_1^{(1)}$, $\bar{\Delta}_1^{(1)}$ respectively. The effect of imposing hypothesis A, under which $\alpha_j=\bar{\alpha}_j$, $\beta_j=\bar{\beta}_j$, $k_j=\bar{k}_j$, is to make $\theta_0=\bar{\theta}_0$, $\theta_1=\bar{\theta}_1$, so that E_j and $\bar{E}_j$ may be viewed as two similar forms involving certain bounded exponentials in y, with coefficients in which the leading asymptotic terms are identical.[6] Hence, under our hypothesis, $E_j-\bar{E}_j=[0]=0(1)/\rho$, and therefore

$$n\rho^{n-1}\left(\frac{\partial G}{\partial x}-\frac{\partial\bar{G}}{\partial x}\right)=0(1)$$

when ρ is on γ'.

The argument is exactly similar for the case ρ on γ''; the only changes required are in the summations of the last line of (3) where the ranges now must be $(1,\mu-1)$, (μ,n). Hence for ρ on γ, and therefore also for ρ on Γ, we have

$$n\rho^{n-1}\left(\frac{\partial G}{\partial x}-\frac{\partial\bar{G}}{\partial x}\right)=0(1),$$

where

---

[5] For a more explicit description of $\Delta_1^{(1)}$ see (S), p. 745 and p. 717.

[6] See (S), p. 729.

from which we obtain

$$\int_{\Gamma} n\rho^{n-1}\left(\frac{\partial G}{\partial x}-\frac{\partial \bar{G}}{\partial x}\right) d\rho = 0(N).$$

This result holds uniformly on $0 \leqq x \leqq 1$.

The case when $n = 2\mu$ may be treated in an entirely analogous manner.

With the help of Lemma 2 we can now establish a theorem concerning the behaviour of $Q(a', b')$ and the related function $\bar{Q}(a', b')$, associated with system (2), when $a' \to 0$ or $b' \to 1$.

THEOREM I. *Under hypothesis A, for fixed b' the functions $Q(a', b')$, $\bar{Q}(a', b')$ are either both bounded as $a' \to 0$, or both become infinite in such a way as to have the same asymptotic behaviour: $Q/\bar{Q} \to 1$ as $a' \to 0$. Similar remarks may be made about the behaviour of Q, $\bar{Q}$ as $b' \to 1$.*

From Lemma 2 we have

$$S'_N - \bar{S}'_N = \int_0^1 S_N(y) \int_{\Gamma} n\rho^{n-1}\left(\frac{\partial G}{\partial x}-\frac{\partial \bar{G}}{\partial x}\right) d\rho dy = 0(N)$$

uniformly on $0 \leqq x \leqq 1$. Hence $S'_N = \bar{S}'_N + 0(N)$, and, the constant Q being the "best" constant, it follows that for some value $x = \xi$ and for some value of N,

$$(Q-1)N \leqq |S'_N(\xi)| = |\bar{S}'_N(\xi) + 0(N)| \leqq \bar{Q}N + |0(N)|.$$

On adding N to both sides and dividing by N we obtain $Q \leqq \bar{Q} + |0(1)|$. Likewise we can show that $\bar{Q} \leqq Q + |0(1)|$. From these results it follows that

$$|Q - \bar{Q}| \leqq K,$$

where K is a constant independent of N and also of a', b'. The conclusions of the theorem are implied in this last result.

Theorem I and Lemma 1 enable us to deduce a useful

COROLLARY. *Under hypothesis A a necessary and sufficient condition that S_N obey Bernstein's theorem on the whole interval $0 \leqq x \leqq 1$ is that $\bar{S}_N$ obey that theorem on the same interval.*

Lemma 1 implies that $S_N - \bar{S}_N = 0(1)$, or, since $|S_N| \leqq 1$, that $|\bar{S}_N| = 0(1)$, uniformly on $0 \leqq x \leqq 1$. On the other hand Theorem I shows that $\bar{Q}$ must be bounded if Q is to be bounded as $a' \to 0$ and $b' \to 1$. But this means that $\bar{S}_N$ obeys Bernstein's theorem on $(0, 1)$.

To illustrate the usefulness of the corollary we shall now give two applications:

I. The Sturm-Liouville case.[7] Let (1) be identified with the Sturm-Liouville system

$$\begin{aligned} &u'' + (\lambda + l(x))u = 0,\\ &u'(0) - hu(0) = 0,\\ &u'(1) + Hu(1) = 0. \end{aligned}$$

and (2) with the system

$$\begin{aligned} &u'' + \lambda u = 0,\\ &u'(0) = u'(1) = 0. \end{aligned}$$

Both systems are normalized and regular and both satisfy the general requirements set forth in the first paragraph. Furthermore, hypothesis A is satisfied, for

$$\begin{matrix} \alpha_1 = \beta_2 = 1, & \beta_1 = \alpha_2 = 0, & k_1 = k_2 = 1,\\ \bar\alpha_1 = \bar\beta_2 = 1, & \bar\beta_1 = \bar\alpha_2 = 0, & \bar k_1 = \bar k_2 = 1. \end{matrix}$$

But system (2) gives rise to sums $\bar S_N$ which are cosine sums on the half period $(0,1)$, and these in turn, being even trigonometric functions, obey Bernstein's theorem on that half period. Hence, by our corollary, there exists a constant Q such that $| S'_N | \leqq QNL$ uniformly on $0 \leqq x \leqq 1$.

II. A general n-th order case. Let (1) be identified with the normalized and regular system

$$\begin{aligned} &L(u) \equiv u^{(n)} + P_2(x)u^{(n-2)} + \cdots + P_n(x)u + \lambda u = 0,\\ &W_j(u) \equiv u^{(n-j)}(0) - u^{(n-j)}(1) + \sum_{i=0}^{n-j-1} (\alpha_{ij}u^{(i)}(0) + \beta_{ij}u^{(i)}(1)) = 0,\\ &\qquad\qquad\qquad\qquad\qquad\qquad (j = 1, 2, \cdots, n), \end{aligned}$$

and (2) with the normalized and regular Fourier system

$$\begin{aligned} &u^{(n)} + \lambda u = 0,\\ &u^{(n-j)}(0) - u^{(n-j)}(1) = 0, \qquad (j = 1, 2, \cdots, n), \end{aligned}$$

both satisfying the further requirement set forth in paragraph 1 concerning the poles of the Green's functions.[8] Hypothesis A is satisfied, for $\alpha_j = \bar\alpha_j = 1$,

---

[7] The Bernstein extension to this case was proved by Miss E. Carlson, using different methods, in a paper: "Extension of Bernstein's theorem to Sturm-Liouville sums," *Transactions of the American Mathematical Society*, vol. 26 (1924), pp. 230-240.

[8] When n is even the characteristic values of the Fourier system appear as double roots of the characteristic equation, but these give rise to simple poles of the Green's function.

$\beta_j = \bar{\beta}_j = -1$, $k_j = \bar{k}_j = n - j$. But the sums $\bar{S}_N$ are trigonometric polynomials on the period interval $(0,1)$, and as such obey Bernstein's theorem on that interval. Hence, in this case also, there exists a constant Q such that $| S'_N | \leqq QNL$ uniformly on $0 \leqq x \leqq 1$.

The importance of Theorem I for the study of the behaviour of $Q(a', b')$ as $a' \to 0$ or $b' \to 1$ lies in the fact that it allows us to carry out the study for the function $\bar{Q}(a', b')$ associated with a simpler type of system than we have in (1). Throughout the rest of the paper we shall identify (2) with the system

$$\begin{aligned} \bar{L}(u) &\equiv u^{(n)} + \lambda u = 0, \\ \bar{W}_j(u) &\equiv \alpha_j u^{(k_j)}(0) + \beta_j u^{(k_j)}(1) = 0, \qquad (j = 1, \cdots, n). \end{aligned}$$

Obviously then hypothesis A is satisfied. We now proceed to obtain an explicit formula for the sum $\bar{S}'_N(x)$ when $x = 0$ (and by analogy when $x = 1$ also). It is necessary to consider separately the cases of odd and even order.

Case 1. $n = 2\mu - 1$. The function $S'_N(x)$ is given by

$$\bar{S}'_N(x) = \frac{1}{2\pi i} \int_0^1 S_N(y) \int_\Gamma n\rho^{n-1} \left\{ \frac{\partial \bar{G}}{\partial x} ; \frac{\partial \bar{G}}{\partial x} \right\} d\rho dy,^{9}$$

where Γ is an arc on the circle $|\rho| = R$, contained on two adjacent sectors which, for definiteness, may be taken to be

$$\begin{aligned} S_1 &: \quad 0 \leqq \arg \rho \leqq \pi/n, \\ S_2 &: \ -\pi/n \leqq \arg \rho \leqq 0. \end{aligned}$$

The two arcs associated with S_1, S_2 we shall denote by γ_1, γ_2, and the two halves of each of these by γ'_1, γ''_1, and γ'_2, γ''_2, respectively. In particular, the arcs γ'_1, γ'_2 are those on which the real part of $\rho\omega_\mu \leqq 0$, whereas the arcs γ''_1, γ''_2 are those on which the real part $\geqq 0$.

From (S, p. 745) we have, on putting $x = 0$ and taking $k = 1$,

$$n\rho^{n-1} \left\{ ; \frac{\partial \bar{G}}{\partial x} \right\} = \left\{ ; F^1{}_{11}(0, y, \rho) + \sum_{j=\mu+1}^{n} e^{-\rho\omega_j y} \frac{m_j(0, y, \rho)}{\rho} \right\} + \frac{\Delta_1{}^{(1)}(0, y, \rho)}{[\theta_0] + e^{\rho\omega_\mu}[\theta_1]} .$$

But, from (M, p. 300),

$$F^1{}_{11}(0, y, \rho) \equiv \sum_{j=\mu+1}^{n} \omega_j e^{-\rho\omega_j y} (A_{11}(x) + B_1(y) + \rho\omega_j),$$

whereas, by using an argument similar to that given in (S, p. 732), we obtain for the last term the formula

---

[9] The notation $\{A ; B\}$ is used to indicate that A is to be taken when $x \geqq y$, and B when $x \leqq y$.

$$\frac{\Delta_1^{(1)}(0, y, \rho)}{[\theta_0] + e^{\rho\omega_\mu}[\theta_1]} = \left(\frac{1}{\theta_0}\delta_1(0, y, \rho) + e^{\rho\omega_\mu}M'(0, y, \rho)\right)\rho$$

where

$$\delta_1(0, y, \rho) = \begin{vmatrix} \omega_1 \cdots \omega_\mu \omega_{\mu+1}e^{-\rho\omega_{\mu+1}} & \cdots \omega_n e^{-\rho\omega_n} & 0 \\ \alpha_1\omega_1^{k_1} \cdots \alpha_1\omega_\mu^{k_1}\beta_1\omega_{\mu+1}^{k_1} & \cdots \beta_1\omega_n^{k_1} & \zeta_1 \\ \vdots \quad \cdots \quad \vdots \quad \vdots & \cdots \quad \vdots & \vdots \\ \alpha_n\omega_1^{k_n} \cdots \alpha_n\omega_\mu^{k_n}\beta_n\omega_{\mu+1}^{k_n} & \cdots \beta_n\omega_n^{k_n} & \zeta_n \end{vmatrix}$$

and

$$\zeta_i(y, \rho) = \sum_{j=1}^{\mu} e^{\rho\omega_j(1-y)}\beta_i\omega_j^{k_i+1} - \sum_{j=\mu+1}^{n} e^{-\rho\omega_j y}\alpha_i\omega_j^{k_i+1}.$$

On substituting these values into the integral

$$I'_1 = \int_0^1 S_N(y) \int_{\gamma'_1} n\rho^{n-1} \left\{ ; \frac{\partial \bar{G}}{\partial x}(0, y, \rho^n) \right\} d\rho dy,$$

and expanding the result sufficiently, we obtain a large number of separate integrals. These latter, however, are all either $0(1)$ or $0(N)$, except those involving $e^{\rho\omega_\mu(1-y)}$. To verify this statement we note that the integrals in question may be identified with the following typical forms:

$$\text{(i)} \quad \int_0^1 S_N(y) \int_{\gamma'_1} \omega_j^2 e^{-\rho\omega_j y}\rho d\rho dy = 0(N), \qquad j = \mu + 1, \cdots, n;$$

$$\text{(ii)} \quad \int_0^1 S_N(y) \int_{\gamma'_1} e^{-\rho\omega_j y}(m_j/\rho) d\rho dy = 0(1), \qquad j = \mu + 1, \cdots, n;$$

$$\text{(iii)} \quad \int_0^1 S_N(y) \int_{\gamma'_1} e^{\rho\omega_\mu}M'\rho d\rho dy = 0(N);$$

$$\text{(iv)} \quad \int_0^1 S_N(y) \int_{\gamma'_1} e^{-\rho\omega_i}m\rho d\rho dy = 0(N), \qquad j = \mu + 1, \cdots, n;$$

$$\text{(v)} \quad \int_0^1 S_N(y) \int_{\gamma'_1} e^{\rho\omega_j(1-y)}\rho d\rho dy = 0(N), \qquad j = 1, \cdots, \mu - 1.$$

(i) may be proved as follows: Let $\rho\omega_j = Re^{i\theta}$; then, when $j = \mu + 1, \cdots, n$, θ will vary over a range (θ_1, θ_2) such that $-\pi/2 < \theta_1, \theta_2 < \pi/2$. Then

$$\left| \int_0^1 S_N(y) \int_{\gamma'_1} \omega_j^2 e^{-\rho\omega_j y}\rho d\rho dy \right| \leqq \int_0^1 \int_{\theta_1}^{\theta_2} R^2 e^{-Ry\cos\theta} d\theta dy$$

$$\leqq R\int_{\theta_1}^{\theta_2} (1 - e^{-R\cos\theta}) \frac{d\theta}{|\cos\theta|} \leqq 2R \int_{\theta_1}^{\theta_2} \frac{d\theta}{|\cos\theta|} = 0(R) = 0(N).$$

(ii) is implied by (S, p. 714, Lemma III). The left-hand member of (iii) may be written

$$\int_0^1 S_N(y) \int_{\gamma'_1} e^{\rho\omega_\mu}M'\rho d\rho dy = R\int_0^1 S_N(y) \int_{\gamma'_1} e^{\rho\omega_\mu}M'' d\rho dy,$$

where $M'' = M'(\rho/R)$, and hence, by (S, p. 732), it is $= R0(1) = 0(N)$. (iv) may be expressed in terms of (iii), since $e^{-\rho\omega_i} = e^{\rho\omega_\mu}0(1)$, $i = \mu + 1, \cdots, n$, and (v) is similar to (i).

As a consequence of the observations which have just been made we see that

$$I'_1 = \frac{B}{\omega_\mu\theta_0}\int_0^1 S_N(y)\int_{\gamma'_1}\omega_\mu{}^2 e^{\rho\omega_\mu(1-y)}\rho d\rho dy + 0(N),$$

where

$$B = \begin{vmatrix} \omega_1, \cdots \omega_\mu & 0 & \cdots 0 & 0 \\ \alpha_1\omega_1{}^{k_1} \cdots \alpha_1\omega_\mu{}^{k_1} & \beta_1\omega_{\mu+1}{}^{k_1} & \cdots \beta_1\omega_n{}^{k_1} & \beta_1\omega_\mu{}^{k_1} \\ \vdots \qquad\qquad \vdots & \vdots & \vdots & \vdots \\ \alpha_n\omega_1{}^{k_n} \cdots \alpha_n\omega_\mu{}^{k_n} & \beta_n\omega_{\mu+1}{}^{k_n} & \cdots \beta_n\omega_n{}^{k_n} & \beta_n\omega_\mu{}^{k_n} \end{vmatrix}$$

and $\theta_0 =$ the minor of the last element of the first row of B.

The corresponding result for the case when ρ is on γ''_1 may be worked in a similar manner. It is found to be

$$I''_1 = \left(1 - \frac{A}{\omega_\mu\theta_1}\right)\int_0^1 S_N(y)\int_{\gamma''_1}\omega_\mu{}^2 e^{-\rho\omega_\mu y}\rho d\rho dy + 0(N),$$

where

$$A = \begin{vmatrix} \omega_1 \cdots \omega_{\mu-1} & 0 & \cdots 0 & 0 \\ \alpha_1\omega_1{}^{k_1} \cdots \alpha_1\omega_{\mu-1}{}^{k_1} & \beta_1\omega_\mu{}^{k_1} & \cdots \beta_1\omega_n{}^{k_1} & \alpha_1\omega_\mu{}^{k_1} \\ \vdots \qquad\qquad \vdots & \vdots & \vdots & \vdots \\ \alpha_n\omega_1{}^{k_n} \cdots \alpha_n\omega_{\mu-1}{}^{k_n} & \beta_n\omega_\mu{}^{k_n} & \cdots \beta_n\omega_n{}^{k_n} & \alpha_n\omega_\mu{}^{k_n} \end{vmatrix}$$

and $\theta_1 =$ the minor of the last element of the first row of A.

Moreover, to carry over these results from the sector S_1 to the sector S_2 it is only necessary to redistribute the subscripts on the ω's. Using "primes" to indicate the results for the sector S_2, we then obtain formulas I'_2, I''_2 exactly identical with those for I'_1, I''_1 except for the replacement of $\omega_1, \omega_2, \cdots, \omega_n$ by $\omega'_1, \omega'_2, \cdots, \omega'_n$. Furthermore, with our choice of the sectors S_1, S_2 it is easily seen that the sequence $\omega'_1, \omega'_2, \cdots, \omega'_n$ is to be identified with the sequence $\omega_1, \omega_3, \omega_2, \omega_5, \omega_4, \cdots, \omega_n, \omega_{n-1}$.

We now observe

LEMMA 3. (a) $\displaystyle\int_{\gamma'_1}\omega_\mu{}^2 e^{\rho\omega_\mu(1-y)}\rho d\rho + \int_{\gamma'_2}\omega'_\mu{}^2 e^{\rho\omega'_\mu(1-y)}\rho d\rho = 0(N)$;

(b) $\displaystyle\int_{\gamma''_1}\omega_\mu{}^2 e^{-\rho\omega_\mu y}\rho d\rho + \int_{\gamma''_2}\omega'_\mu{}^2 e^{-\rho\omega'_\mu y}\rho d\rho = 0(N)$.

It will be sufficient to prove part (a). Consider the two arcs on the circle $|z| = R$, in the complex z-plane, defined as follows:

$$c_1 : \pi/2 \leqq \arg z \leqq \pi/2 + \phi,$$
$$c_2 : 3\pi/2 - \phi \leqq \arg z \leqq 3\pi/2,$$

where $\phi = \pi/(2n)$. The two integrals in (a) above may now be written

$$\int_{c_1} ze^{z(1-y)}dz + \int_{c_2} ze^{z(1-y)}dz.$$

Let c_3 be the arc of the circle $|z| = R$ defined by $\pi/2 + \phi \leqq \arg z \leqq 3\pi/2 - \phi$, and let c_4 be that diameter of the circle which coincides with imaginary axis. Then, on applying Cauchy's integral theorem, we have

$$\int_{c_1} ze^{z(1-y)}dz + \int_{c_2} ze^{z(1-y)}dz = -\int_{c_3} ze^{z(1-y)}dz - \int_{c_4} ze^{z(1-y)}dz.$$

But

$$\int_{c_3} ze^{z(1-y)}dz = \frac{z_1e^{z_1(1-y)} - z_2e^{z_2(1-y)}}{1-y} - \frac{e^{z_1(1-y)} - e^{z_2(1-y)}}{(1-y)^2}$$
$$= \frac{2iR\cos[2R(1-y)\cos\phi + \phi]e^{-R(1-y)\sin\phi}}{1-y} - \frac{2iR\sin[R(1-y)\cos\phi]e^{-R(1-y)\sin\phi}}{(1-y)^2},$$

where $z_1 = R(-\sin\phi + i\cos\phi)$, $z_2 = R(-\sin\phi - i\cos\phi)$ are the extremities of the arc c_3. When $y \neq 1$ this expression is easily seen to be $0(R) = 0(N)$ as $R \to \infty$, whereas when $y \to 1$ the expression converges to zero for any given value of R. Hence

$$\int_{c_3} ze^{z(1-y)}dz = 0(N).$$

Similarly,

$$\int_{c_4} ze^{z(1-y)}dz = \int_{-R}^{R} (it)e^{it(1-y)}(idt)$$
$$= \frac{-2R\cos R(1-y)}{1-y} + \frac{2\sin R(1-y)}{(1-y)^2} = 0(N),$$

and thus part (a) of the lemma is established.

On adding the integrals I'_1, I''_1, I'_2, I''_2, and eliminating the integrals involving ω'_μ with the help of Lemma 3, we obtain

$$(5) \qquad 2\pi i\bar{S}'_N(0) = \left(\frac{B}{\omega_\mu\theta_0} - \frac{B'}{\omega'_\mu\theta'_0}\right)\int_0^1 S_N(y) \int_{\gamma'_1} \omega_\mu{}^2 e^{\rho\omega_\mu(1-y)}\rho d\rho dy$$
$$- \left(\frac{A}{\omega_\mu\theta_1} - \frac{A'}{\omega'_\mu\theta'_1}\right)\int_0^1 S_N(y) \int_{\gamma''_1} \omega_\mu{}^2 e^{-\rho\omega_\mu y}\rho d\rho dy + 0(N).$$

Case 2. $n = 2\mu$. The treatment in this case is entirely analogous to that of the foregoing case. We shall merely state the results:

$$(6)\qquad 2\pi i\bar{S}'_N(0) = \left(\frac{B}{\omega_\mu\theta_2} - \frac{B'}{\omega'_\mu\theta'_2}\right)\int_0^1 S_N(y)\int_{\gamma_1}\omega_\mu{}^2 e^{\rho\omega_\mu(1-y)}\rho\,d\rho\,dy$$
$$-\left(\frac{A}{\omega_{\mu+1}\theta_2} - \frac{A'}{\omega'_{\mu+1}\theta'_2}\right)\int_0^1 S_N(y)\int_{\gamma_2}\omega^2{}_{\mu+1}e^{-\rho\omega_{\mu+1}y}\rho\,d\rho\,dy + 0(N),$$

where B is identical with the determinant B of case 1, θ_2 is identical with θ_0 of case 1, A is the same as B except that in the last column the elements are $0, \alpha_1\omega^{k_1}{}_{\mu+1}, \cdots, \alpha_n\omega^{k_n}{}_{\mu+1}$, and B', A' and θ'_2 are the corresponding forms involving $\omega'_1, \omega'_2, \cdots, \omega'_n$.

The integrals remaining in (5) and (6) are definitely $0(N^2)$. Hence, for a fixed value of $b' < 1$, the functions $\bar{Q}(a', b')$, $Q(a', b')$ are bounded as $a' \to 0$ if, and only if, the coefficients of these integrals vanish. Thus we are led to state

THEOREM II. *The necessary and sufficient conditions under which the sums $S_N(x)$ associated with system* (1) *obey Bernstein's theorem on the interval $0 \leqq x \leqq b' < 1$ are as follows:*

(i) *when* $n = 2\mu - 1$,

$$\frac{B}{\omega_\mu\theta_0} - \frac{B'}{\omega'_\mu\theta'_0} = 0, \qquad \frac{A}{\omega_\mu\theta_1} - \frac{A'}{\omega'_\mu\theta'_1} = 0;$$

(ii) *when* $n = 2\mu$,

$$\frac{B}{\omega_\mu\theta_2} - \frac{B'}{\omega'_\mu\theta'_2} = 0, \qquad \frac{A}{\omega_{\mu+1}\theta_2} - \frac{A'}{\omega'_{\mu+1}\theta'_2} = 0.$$

Analogous results may be worked out for the case when $x = 1$.

Theorem II may be applied directly to the Fourier or Sturm-Liouville systems to establish the extension of the Bernstein theorem to the point $x = 0$. On the other hand, in the case of the system $u''' + \lambda u = 0$, $u''(0) = 0$, $u''(1) = 0$, $u'(0) + u'(1) = 0$, for example, the theorem shows that the extension to the point $x = 0$ is not possible.

MOUNT ALLISON UNIVERSITY,
SACKVILLE, N. B., CANADA.

ON THE LINEARITY OF PENCILS OF CURVES ON ALGEBRAIC SURFACES.*

By O. F. G. Schilling[1] and O. Zariski.

The object of this note is to give an arithmetical proof of the following often used theorem: "*If a pencil of curves on an algebraic surface has a base point at a simple point of the surface then the pencil is either a linear system or its curves are cut out by hypersurfaces* $(\phi + \lambda\psi)^{\rho}$." The essential feature of our approach consists in eliminating the difficulties which arise from the possible singularities of the curves at the simple base point. The interpretation of the pencil of curves as a rational transform of the given surface allows us to apply a theorem proved by one of us.[2]

Let K be a field of algebraic functions of two variables over an algebraically closed field k. An algebraic surface f in the affine n-dimensional space S_n over k is said to be a model of K in S_n if the quotient field of $k[x_1, \cdots, x_n]/\mathfrak{p}(f)$ which is determined by the prime ideal $\mathfrak{p}(f)$ defining the surface f, is isomorphic with K. Thus f is described by the order $k[\xi_1, \cdots, \xi_n] = \mathfrak{O}$ in K where $\xi_i = x_i \bmod \mathfrak{p}(f)$. The 0-dimensional prime ideals $\mathfrak{p}$ of $k[x_1, \cdots, x_n]$ which divide $\mathfrak{p}(f)$ correspond to the points P of the surface f. A point P with the coördinates $\{a_1, \cdots, a_n\}$ is called a simple point of f if

(i) the ideal $(\xi_1 - a_1, \cdots, \xi_n - a_n)$ is a 0-dimensional prime ideal $\mathfrak{p}$ in the integral closure of $\mathfrak{O}$ and

(ii) it is possible to choose two algebraically independent elements, say ξ_1, ξ_2 among $\xi_1, \xi_2, \cdots, \xi_n$ such that the ideal $(\xi_1 - a_1, \xi_2 - a_2)$ is divisible by $\mathfrak{p}$ but not divisible by any primary ideal belonging to $\mathfrak{p}$.

It can be shown that all elements of $\mathfrak{O}$ can be expanded in formal power series of $u = \xi_1 - a_1$, $v = \xi_2 - a_2$ with coefficients in k.[3] Hence the elements of $\mathfrak{O}$ are contained in the ring of holomorphic functions $\{\sum_{i,j \geqq 0} a_{ij} u^i v^j\}$ which itself is contained in the field of all formal meromorphic functions of u, v:

---

* Received February 7, 1938.

[1] Johnston Scholar of the Johns Hopkins University for 1937-1938.

[2] O. Zariski, "Polynomial ideals defined by infinitely near base points," § 13, *American Journal of Mathematics*, vol. 60 (1938).

[3] O. Zariski, "Some results in the arithmetic theory of algebraic functions of several variables," *Proceedings of the National Academy of Sciences*, vol. 23 (1937).

$$k\{u, v\} = \{(\Sigma a_{ij} u^i v^j)(\Sigma b_{ij} u^i v^j)^{-1}\}.$$

An irreducible algebraic system Σ_r of curves on the surface f is given by an irreducible algebraic correspondence between f and a r-dimensional algebraic variety V_r such that to a generic point on V_r there corresponds a curve $C \subset \Sigma_r$ on f. A pencil Σ_1 of curves on f is an irreducible algebraic system Σ_1 such that there passes through a generic point P of f exactly one curve C in Σ_1.[4] This definition of a pencil Σ_1 is equivalent to the following: the function field K_1 belonging to the variety V_1 defining Σ_1 is a rational transform of the surface f, i. e. K_1 is isomorphic with a 1-dimensional subfield $\bar{K}_1$ of K.

If V_r is a linear r-dimensional space and if the curves of Σ_r are cut out by hypersurfaces $\phi = \sum_{i=0}^{r} \lambda_i \phi_i = 0$, ϕ_i being forms in the imbedding space of f, then Σ_r is called a linear system. In a linear system one usually omits the fixed curves which are cut out by all hypersurfaces ϕ.

After these preliminary remarks we proceed to the proof of the

THEOREM. *If a pencil of curves Σ_1 on an algebraic surface f has a base point at a simple point of f then Σ_1 is either a linear system or its curves are cut out by hypersurfaces $(\phi + \lambda\psi)^p = 0$.*

Proof. Let P be the simple base point of the pencil Σ_1. Since P is assumed to be a simple point of the surface f, there exist functions u, v in $\mathfrak{O}$ defining a field of meromorphic functions $k\{u, v\}$ which contains a subfield $K^* \cong K$. Moreover, $u = v = 0$ at P. Consequently, the field $\bar{K}_1$ which belongs to the pencil Σ_1 has also an isomorphic map K^*_1 in $k\{u, v\}$. Thus each element $a^* \varepsilon K^*$ is represented by a ratio $\dfrac{\alpha(u, v)}{\beta(u, v)}$ of holomorphic functions $\alpha(u, v)$, $\beta(u, v)$. Moreover, there exists a function $\dfrac{\alpha(u, v)}{\beta(u, v)} = A^* \varepsilon K^*_1$ such that

$$\alpha(0, 0) = \beta(0, 0) = 0$$

when u and v assume the constant values and $0, 0$, respectively, at the given point P. The existence of such a function A^* is a consequence of the assumption that P is a simple base point of Σ_1, i. e. that there corresponds to P the whole curve V_1 under the correspondence between f and V_1. In fact, let us assume that the surface f is given in a 3-dimensional affine space $k[x_1, x_2, x_3]$ and that V_1 is given in an n-dimensional projective space $k[y_0, y_1, \cdots, y_n]$. Since K_1 is a rational transform of K we have relations of the following type

---

[4] For these definitions see for example O. Zariski, "Algebraic surfaces," Chapters II, V, *Ergebnisse der Mathematik und ihrer Grenzgebiete* (Berlin, 1935).

$$P(x_1, x_2, x_3)y_i - Q_i(x_1, x_2, x_3)y_0 = 0 \qquad (i = 1, 2, \cdots, n)$$

where $P(x_1, x_2, x_3)$ and $Q_i(x_1, x_2, x_3)$ are polynomials in x_1, x_2, x_3. Using the imbedding of K as K^* in $k\{u, v\}$ we obtain

$$P_i(u, v)y_i - Q_i(u, v)y_0 = 0 \qquad (i = 1, 2, \cdots, n)$$

where $P_i(u, v)$ and $Q_i(u, v)$ are relatively prime holomorphic functions in u, v. These equations can be considered as relations which are contained in the ideal $\mathfrak{C}$ of relations defining the correspondence. The assumption that P be a base point of Σ_1 implies then that

$$P_i(0, 0)y_i - Q_i(0, 0)y_0 \equiv 0 \text{ in } K^*_1.$$

We may suppose that y_0 is different from 0, then

$$P_i(0, 0)\frac{y_i}{y_0} - Q_i(0, 0) \equiv 0.$$

Consequently, since at least one function y_i/y_0 of the field $K^*_1 \cong K_1$ does not lie in k,

$$P_i(0, 0) = Q_i(0, 0) = 0,$$

or y_i/y_0 is a function having the desired properties. Now we are in a position to apply a result of the general theory of valuation ideals stating that for each function $A^* = \dfrac{\alpha(u, v)}{\beta(u, v)}$, where $\alpha(0, 0) = \beta(0, 0) = 0$, and α, β are relatively prime, there exists a prime divisor $\mathfrak{P}$ of $k\{u, v\}$ which maps $k\{u, v\}$ upon a purely transcendental field $k(t)$ in which the map $A^* * \mathfrak{P}$ of A^* is a transcendental quantity with respect to k.[5] Consequently, the field $K^*_1 \cong K_1$ is mapped upon a transcendental subfield of $k(t)$. Hence K^*_1 is itself a purely transcendental subfield, for the divisor $\mathfrak{P}$ acts as an isomorphic mapping on K^*_1, since K^*_1 and its map have the same degree of transcendentality. We have

$$K^*_1 \cong \bar{K}_1 = k(\lambda) \subset K$$

where λ is the ratio $\dfrac{p(x_1, x_2, x_3)}{q(x_1, x_2, x_3)}$ of relatively prime polynomials $p(x_1, x_2, x_3)$ and $q(x_1, x_2, x_3)$ in $k[x_1, x_2, x_3]$.

We observe that we do not change the nature of the algebraic pencil Σ_1 if we use instead of the original variety V_1 the birationally equivalent curve $k[\lambda]$ for the definition of Σ_1.

Now it remains to be shown that Σ_1 is a linear system cut out by surfaces

---

[5] See note 2, *loc. cit.*, p. 203

$\lambda q(x_1, x_2, x_3) - p(x_1, x_2, x_3) = 0$ or a system of curves cut out by the surfaces $(\lambda q(x_1, x_2, x_3) - p(x_1, x_2, x_3))^\rho = 0$, $\rho > 1$. Consider for this purpose the ideal

$$\mathfrak{A} = (\lambda(x_1, x_2, x_3) - p(x_1, x_2, x_3), f) = (\lambda p - q, f)$$

in the ring $k[x_1, x_2, x_3, \lambda]$ where $(f) = \mathfrak{p}(f)$ denotes the prime ideal defining the surface f. According to a well-known theorem of Macauly the ideal $\mathfrak{A}$ is unmixed of dimension 2, thus

$$\mathfrak{A} = [\mathfrak{q}_1, \mathfrak{q}_2, \cdots, \mathfrak{q}_s]$$

where the ideals $\mathfrak{q}_i$ are 2-dimensional primary ideals with the associated prime ideals $\mathfrak{P}_i$. The contracted ideal $\bar{\mathfrak{A}} = \mathfrak{A} \cap k[x_1, x_2, x_3]$ of $\mathfrak{A}$ is equal to (f). Hence

$$\bar{\mathfrak{A}} = (f) = [\bar{\mathfrak{q}}_1, \bar{\mathfrak{q}}_2, \cdots, \bar{\mathfrak{q}}_s]$$

where $\bar{\mathfrak{q}}_i = \mathfrak{q}_i \cap k[x_1, x_2, x_3]$. This representation implies that one component $\bar{\mathfrak{q}}_i$ must be equal to (f); let $\bar{\mathfrak{q}}_1$ be such a component.

We consider next an arbitrary element $F(x_1, x_2, x_3, \lambda)$ lying in $\mathfrak{q}_1$, then

$$q^\sigma F(x_1, x_2, x_3, \lambda) = A(x_1, x_2, x_3, \lambda)(\lambda q - p) + B(x_1, x_2, x_3).$$

Since $F(x_1, x_2, x_3, \lambda)$ and $\lambda p - q$ both lie in $\mathfrak{q}_1$ we get

$$B(x_1, x_2, x_3) \subset \mathfrak{q}_1,$$

consequently

$$\begin{aligned} B(x_1, x_2, x_3) &\equiv C \pmod{\bar{\mathfrak{q}}_1} \\ &\equiv C \pmod{f}. \end{aligned}$$

There therefore exists a common exponent $\tau > 0$ such that

$$q^\tau F(x_1, x_2, x_3, \lambda) \equiv 0 \pmod{\mathfrak{A}}$$

for any element $F \subset \mathfrak{q}_1$, because $\mathfrak{q}_1$ has a finite base. Hence

$$q^\tau \mathfrak{q}_1 \equiv 0 \pmod{\mathfrak{q}_2},$$

consequently

$$q^\tau \equiv 0 \pmod{\mathfrak{q}_2},$$

for $\mathfrak{A}$ is an unmixed ideal and consequently all components $\mathfrak{q}_i$ have the same dimension. Therefore

$$q \equiv 0 \pmod{\mathfrak{P}_2}$$

and also

$$q \equiv 0 \ (\mathrm{mod}\ \bar{\mathfrak{P}}_2).$$

Since $\lambda = p/q$ we have $q \not\equiv 0(f)$, thus

$$\bar{\mathfrak{q}}_2 \not\equiv 0(f) \quad \text{and} \quad \bar{\mathfrak{P}}_2 \not\equiv 0(f).$$

Therefore $\bar{\mathfrak{q}}_2 \supset (f)$, consequently the ideals $\bar{\mathfrak{q}}_2, \cdots, \bar{\mathfrak{q}}_3$ are 0- or 1-dimensional ideals. They must be 1-dimensional, for $\bar{\mathfrak{q}}_2 k[x_1, x_2, x_3, \lambda] \subset \mathfrak{q}_2$ and hence

$$\begin{aligned} \dim \mathfrak{q}_2 = 2 &\leq \dim \bar{\mathfrak{q}}_2 k[x_1, x_2, x_3, \lambda] \\ &= \dim \bar{\mathfrak{q}}_2 + 1, \end{aligned}$$

or

$$\dim \bar{\mathfrak{q}}_2 = 1.$$

This relation between the dimensions of $\mathfrak{q}_2$ and $\bar{\mathfrak{q}}_2$ shows that the components $\bar{\mathfrak{q}}_2, \cdots, \bar{\mathfrak{q}}_s$ do not depend on λ, i. e. they are extended ideals of $\bar{\mathfrak{q}}_2, \cdots, \bar{\mathfrak{q}}_s$. In geometric terms, the curves $\mathfrak{q}_2, \cdots, \mathfrak{q}_s$ correspond to the entire line $k[\lambda]$ under the algebraic correspondence. Such fixed components shall be left out in the definition of a linear system, and consequently $\mathfrak{q}_1 = \mathfrak{q}$ is the ideal defining Σ_1. According to the properties of primary ideals we have

$$(\lambda q - p)^{\rho} \subset \mathfrak{q},$$

i. e. Σ_1 is cut out by the hypersurfaces $(\lambda q - p)^{\rho} = 0$ where fixed components are omitted.

We remark that if q lies in one of the components $\mathfrak{q}_i \neq \mathfrak{q}$ then also $p \subset \mathfrak{q}_i$ for $\lambda q - p \subset \mathfrak{A}$; hence $p = q = 0$ occur among the equations defining the correspondence $\mathfrak{q}$.

THE JOHNS HOPKINS UNIVERSITY.

SOME SINGULAR PROPERTIES OF CONFORMAL TRANSFORMATIONS BETWEEN RIEMANNIAN SPACES.*

By VIRGINIA MODESITT.

1. Introduction. Two Riemannian spaces V_n and V'_n are in conformal correspondence if their fundamental tensors are related by[1]

$$g'_{ij} = e^{2\sigma} g_{ij} \qquad (i, j = 1, \cdots n), \tag{1.1}$$

where σ is any function of the x's.[2] The purpose of this paper is to investigate some geometric properties of corresponding curves and subspaces in V_n and V'_n.

From (1.1), it follows that

$$g'^{ij} = e^{-2\sigma} g^{ij}; \tag{1.2}$$

that contravariant and covariant components of corresponding unit vectors are related by

$$\lambda'^i = e^{-\sigma}\lambda^i, \qquad \lambda'_i = e^{\sigma}\lambda_i; \tag{1.3}$$

and that Christoffel symbols for the two spaces are given by

$$\begin{aligned} [ij, k]' &= e^{2\sigma}([ij, k] + g_{ik}\bar{\sigma}_j + g_{jk}\bar{\sigma}_i - g_{ij}\bar{\sigma}_k), \\ \left\{ {i \atop jk} \right\}' &= \left\{ {i \atop jk} \right\} + \delta_j{}^i\bar{\sigma}_k + \delta_k{}^i\bar{\sigma}_j - g_{jk}\bar{\sigma}^i, \end{aligned} \tag{1.4}$$

where $\bar{\sigma}_i = \partial\sigma/\partial x^i$, $\bar{\sigma}^i = g^{im}\bar{\sigma}_m$. The bar is used throughout to indicate that the components of a vector are not necessarily unit. It is to be noted that σ^i are the components of the congruence of curves normal to the family of hypersurfaces, $\sigma =$ constant. We shall call this congruence the congruence of σ-curves and the family of hypersurfaces H_σ.

It will also be useful to have the relation between the covariant derivatives of unit contravariant components of corresponding directions. From (1.3) and (1.4), we obtain

$$\lambda'^i{}_{,j} = e^{-\sigma}(\lambda^i{}_{,j} + \delta_j{}^i\bar{\sigma}_k\lambda^k - g_{jk}\lambda^k\bar{\sigma}^i), \tag{1.5}$$

---

* Received July 17, 1937.

[1] L. P. Eisenhart, *Riemannian Geometry*, Princeton University Press, 1926, p. 89. We shall refer to this book as R. G.

[2] We shall assume in what follows that the fundamental forms of V_n and V'_n are positive definite and that the function σ is not identically a constant.

where the dot is used to denote covariant differentiation with respect to the g''s and the comma to denote covariant differentiation with respect to the g's.

2. Curves with corresponding principal normals. The p-th normals of a curve C are given by the Frenet formulas[3]

$$(2.1) \qquad {}_{p+1}\lambda^i = \rho_p({}_1\lambda^j\, {}_p\lambda^i{}_{,j} + (1/\rho_{p-1})\, {}_{p-1}\lambda^i) \qquad \begin{matrix}(p = 1, \cdots, n-1)\\ (1/\rho_0 = 1/\rho_n = 0),\end{matrix}$$

where the p-th curvatures of C satisfy the conditions

$$(2.2) \qquad 1/\rho_p = {}^{p+1}\lambda_i\, {}_1\lambda^j\, {}_p\lambda^i{}_{,j}. \qquad (p = 1, \cdots, n-1).$$

Similar relations in the primed quantities may be written for the normals and curvatures of the corresponding curve C'.

In particular, the principal normals of C' are given by

$$(2.3) \qquad {}_2\lambda'^i = \rho'_1\, {}_1\lambda'^j\, {}_1\lambda'^i{}_{.j}.$$

Since the tangent vectors of C and C' are in corresponding directions, their unit components are related by ${}_1\lambda'^i = e^{-\sigma}\, {}_1\lambda^i$. By means of (1.5), it follows from (2.3) that principal normals of C and C' are in the relation

$$(2.4) \qquad {}_2\lambda'^i/\rho'_1 = e^{-2\sigma}({}_2\lambda^i/\rho_1 + \bar{\sigma}_k\, {}_1\lambda^k\, {}_1\lambda^i - \bar{\sigma}^i).$$

Hence, in general, the principal normals of C and C' are not in corresponding directions. If their directions do correspond, the vector, $\bar{\sigma}_k\, {}_1\lambda^k\, {}_1\lambda^i - \bar{\sigma}^i$, must either have zero components or else must be in the direction ${}_2\lambda^i$. In the first case, C is a $\bar{\sigma}$-curve. We shall discuss such curves more particularly in the next section and confine ourselves here to the second case,—that in which we can write

$$(2.5) \qquad {}_2\lambda^i = a\, {}_1\lambda^i + b\bar{\sigma}^i.$$

Let us consider the surface generated by the σ-curves at points of a curve and call the V_2 so formed the σ-surface of the curve. From (2.5), it is seen that for a curve C of the type we are considering, the principal normals are directions in the σ-surface of C, and conversely, if ${}_2\lambda^i$ is a direction in the σ-surface of C, then C and C' are curves with corresponding principal normals.

The principal normals of a curve (not geodesic in V_n) which lies in a subspace V_m of V_n, $(m < n)$, may be written[4]

$$(2.6) \qquad {}_2\lambda^i/\rho_1 = \nu^i/\rho_g + \zeta^i/R$$

---

[3] E. G., p. 106.

[4] E. G., p. 165.

where ν^i are the unit components of the principal normal of the curve in V_m, and ζ^i is the normal curvature vector to V_m along C. $1/\rho_g$ is the first curvature of C in V_m, and $1/R$ is the normal curvature of V_m for the given curve.

Since, for a curve with corresponding principal normals, the directions of these principal normals coincide with the directions of the normals to C in the σ-surface of C, it follows that $1/R$ must be zero, i. e., that such a curve be asymptotic on its σ-surface. Conversely, if V_2 be any surface generated by a one-parameter family of σ-curves, and if C be asymptotic on this surface, then from (2. 6) it is seen that ${}_2\lambda^i$ is a direction in the V_2 and hence that C and C' have corresponding principal normals.

If C is geodesic in V_m, it is both geodesic and asymptotic on its σ-surface, V_2. It is found from (2. 4) that the principal normals of C' are directions in V'_2 and hence that C' is also asymptotic on its σ-surface. In this case, however, the principal normals of C are indeterminate.

We may now state the theorem: *If the principal normals of C and C' are in corresponding directions, (C is not a σ-curve and is not geodesic in V_n), then C is asymptotic on its σ-surface. Conversely, if C, (not a σ-curve and not geodesic in V_n), is asymptotic on its σ-surface, then the principal normals of C and C' are in corresponding directions.*

Let us assume that any two consecutive normals of two corresponding curves C and C', say the $(p-2)$-nd and $(p-1)$-st, $(p > 2)$, are in corresponding directions:

$$ {}_{p-1}\lambda'^i = e^{-\sigma}\,{}_{p-1}\lambda^i, \qquad {}_{p}\lambda'^i = e^{-\sigma}\,{}_{p}\lambda^i \qquad (p > 2). \tag{2. 7}$$

If we write the formula for the $(p-1)$-st curvature of C' from (2. 2) and replace the primed quantities in terms of the unprimed on the right hand side, we will obtain

$$ 1/\rho'_{p-1} = e^{-\sigma}(1/\rho_{p-1}). \tag{2. 8}$$

From (2. 7), (2. 8), (1. 5), and the Frenet formula for the p-th normal of C', we obtain

$$ {}_{p+1}\lambda'^i = e^{-2\sigma}\rho'_p\big((1/\rho_p)\,{}_{p+1}\lambda^i + \bar{\sigma}_k\,{}_p\lambda^k\,{}_1\lambda^i\big). $$

Multiplying both sides by ${}^1\lambda_i$, summing on i, gives

$$ \bar{\sigma}_k\,{}_p\lambda^k = 0 \qquad (p > 2), \tag{2. 9}$$

whence it follows that

$$ {}_{p+1}\lambda'^i = e^{-\sigma}\,{}_{p+1}\lambda^i, \qquad \rho'_p/\rho_p = e^{\sigma}. \tag{2. 10}$$

If, then, any two consecutive normals, ${}_{p-1}\lambda^i$ and ${}_p\lambda^i$, $(p > 2)$, of C and C'

are in corresponding directions, all succeeding normals will also be in corresponding directions and the ratios, ρ'_m/ρ_m, of the m-th radii of curvature of the two curves, $(m = p - 1, \cdots, n)$, will be equal to the coefficient of magnification, e^σ.

Similarly, it can be shown that if the p-th and $(p-1)$-st normals of C and C' are in corresponding directions, then all normals before the $(p-1)$-st are also in corresponding directions and the radii of curvature of the two curves before the p-th will be in the ratio e^σ, with the exception of the radii of first curvature. Since, in particular, we will have ${}_2\lambda'^i = e^{-\sigma}\,{}_2\lambda^i$, C and C' are curves with corresponding principal normals. It follows from (2.2) that, for such curves, first curvatures are in the relation

$$1/\rho'_1 = e^{-\sigma}(1/\rho_1 - \bar{\sigma}_k\,{}_2\lambda^k). \tag{2.11}$$

If the normal ennuple of C, ${}_1\lambda^i, {}_2\lambda^i, \cdots, {}_n\lambda^i$, be considered as a cyclical set of directions, and hence ${}_n\lambda^i$ and ${}_1\lambda^i$, ${}_1\lambda^i$ and ${}_2\lambda^i$, as pairs of consecutive directions, the above results still hold. Consequently: *If any two consecutive directions of the normal ennuples of C and C' correspond, where the directions, ${}_1\lambda^i, \cdots, {}_n\lambda^i$, are considered as a cyclical set, then the remaining directions for the two curves also correspond, and the ratios, ρ'_p/ρ_p, $(p = 2, \cdots, n-1)$, of the p-th radii of curvature are equal to the coefficient of magnification, e^σ.*

3. Properties of σ-curves. We have defined σ-curves as the congruence of curves orthogonal to the hypersurfaces H_σ. If C is a σ-curve,

$$\bar{\sigma}^i = \tau\,{}_1\lambda^i, \qquad \bar{\sigma}_i = \tau\,{}^1\lambda_i. \tag{3.1}$$

Hence, it follows from (2.4) that the σ-curves are curves with corresponding principal normals. Equation (2.11) shows that they differ from the curves with corresponding principal normals already considered in that first curvatures are related by

$$1/\rho'_1 = e^{-\sigma}(1/\rho_1). \tag{3.2}$$

Since the σ-curves fulfill the conditions of the last theorem of section two, the ratios of the p-th radii of curvature of C and C' are also equal to e^σ, $(p = 2, \cdots, n-1)$, and the p-th normals, $(p = 2, \cdots, n)$ are in corresponding directions.

Conversely, if C and C' have corresponding principal normals, we may write, $\bar{\sigma}^i = \alpha\,{}_1\lambda^i + \beta\,{}_2\lambda^i$, and if $1/\rho'_1 = e^{-\sigma}(1/\rho_1)$, it follows that $\beta = 0$ and C is a σ-curve. Hence: *Necessary and sufficient conditions that C be a σ-curve are 1) that C and C' have corresponding principal normals, and 2) that the ratio of the radii of first curvature of C and C' be equal to the coefficient of magnification, e^σ.*

If directions μ^i are parallel along a curve C, we have

$$\mu^i_{,j}\,{}_1\lambda^j = 0. \tag{3.3}$$

Writing the same condition for corresponding directions in V'_n, we find by means of (1.5) that

$$\mu'^i_{.j}\,{}_1\lambda'^j = e^{-2\sigma}(\mu^i_{,j}\,{}_1\lambda^j + \bar{\sigma}_k\mu^k\,{}_1\lambda^i - g_{jk}\,{}_1\lambda^j\mu^k\bar{\sigma}^i). \tag{3.4}$$

It follows from (3.1) and (3.3) that *if directions μ^i are parallel along a σ-curve C, the corresponding directions are parallel along C'.* In particular, if C is geodesic in V_n, from (3.4) for $\mu^i = {}_1\lambda^i$ we may state: *A necessary and sufficient condition that a curve C, geodesic in V_n, correspond to a geodesic in V'_n is that C be a σ-curve.*

4. Osculating spaces of a curve. By the osculating space, O_p, of a curve is meant the V_p determined at a point by the first p directions of the normal ennuple of C. As a special case, it appears, from (2.4), that O'_2 will correspond to O_2 if and only if C is a curve with corresponding principal normals.

It can be shown by induction that ${}_p\lambda'$ can be written as a linear combination of the directions, ${}_1\lambda^i, \cdots, {}_p\lambda^i$, σ^i, ${}_1\eta^i, \cdots, {}_{p-2}\eta^i$, where ${}_k\bar{\eta}^i = {}_{k-1}\eta^i_{,j}\,{}_1\lambda^j$, $(k = 2, \cdots, p-2)$, is the k-th associate direction of σ^i along C. If σ^i, ${}_1\eta^i, \cdots, {}_{p-2}\eta^i$ are directions in the O_p of C, then O_p and O'_p correspond. The converse is also true. We may write the equations

$$\begin{aligned}
{}_2\lambda'^i &= {}_2a\,{}_2\lambda^i + {}_2b\,{}_1\lambda^i + {}_2c\sigma^i,\\
{}_3\lambda'^i &= {}_3a\,{}_3\lambda^i + {}_3b\,{}_2\lambda^i + {}_3c\,{}_1\lambda^i + {}_3d\sigma^i + {}_3e\,{}_1\eta^i,\\
&\cdots\cdots\cdots\cdots\cdots\cdots\\
{}_p\lambda'^i &= {}_pa\,{}_p\lambda^i + \cdots + {}_pb\,{}_1\lambda^i + {}_pc\sigma^i + {}_pd\,{}_1\eta^i + \cdots + {}_pe\,{}_{p-2}\eta^i.
\end{aligned} \tag{4.1}$$

If any ${}_k\eta^i$ is a zero vector, then ${}_r\eta^i$, $(r > k)$, are also zero vectors. For $r < k$, the e's are not zero and under the hypothesis that O_p and O'_p correspond, equations (4.1) can be solved for σ^i, ${}_1\eta^i, \cdots, {}_{k-1}\eta^i$ as linear functions of ${}_1\lambda^i, \cdots, {}_p\lambda^i$. Hence: *If σ^i and its $(p-2)$ associate directions (which may or may not be zero vectors) lie in O_p, the osculating spaces of C and C' correspond and conversely.*

5. Properties of curves in corresponding subspaces. Let C and C' be curves in subspaces V_m and V'_m, $(m < n)$, immersed in V_n and V'_n.[5] We shall denote by ${}_\tau\xi^i$, $(\tau = 1, \cdots, n-m)$, a set of $(n-m)$ mutually orthogonal unit directions in V_n normal to V_m. Corresponding to each of these normals a tensor ${}^{\tau}\Omega_{\alpha\beta}$ is defined by [6]

---

[5] R. G., p. 143 ff., gives a detailed account of the geometry of subspaces.

$$(5.1) \qquad {}^{\tau}\Omega_{\alpha\beta} = g_{ij}x^i{}_{,\alpha\beta}\, {}_{\tau}\xi^j + [kl, j]_g\, x^k{}_{,\alpha} x^l{}_{,\beta}\, {}_{\tau}\xi^j \qquad (\alpha, \beta = 1 \cdots m)$$

and is related to the corresponding tensor ${}^{\tau}\Omega'_{\alpha\beta}$ formed for the direction ${}_{\tau}\xi'^i$ by

$$(5.2) \qquad {}^{\tau}\Omega'_{\alpha\beta} = e^{\sigma}\, ({}^{\tau}\Omega_{\alpha\beta} - a_{\alpha\beta}\bar{\sigma}_k\, {}_{\tau}\xi^k),$$

where $a_{\alpha\beta}$ are the coefficients of the first fundamental form of V_m.

The normal curvature vector of V_m in a given direction is defined by the relation

$$(5.3) \qquad \zeta^i = {}^{\tau}\Omega_{\alpha\beta}\, {}_1\lambda^{\alpha}\, {}_1\lambda^{\beta}\, {}_{\tau}\xi^i$$

and is related to the normal curvature vector of V'_m by

$$(5.4) \qquad \zeta'^i/R' = e^{-2\sigma}(\zeta^i/R - \sum_{\tau=1}^{n-m} \bar{\sigma}_k\, {}_{\tau}\xi^k\, {}_{\tau}\xi^i),$$

where $1/R$ is the normal curvature of the V_m for the given direction. Since $\sum_{\tau=1}^{n} \bar{\sigma}_k\, {}_{\tau}\xi^k\, {}_{\tau}\xi^i = \bar{\sigma}_k g^{ki} = \bar{\sigma}^i$, we may also write (5.4) in the form

$$(5.5) \qquad \zeta'^i/R' = e^{-2\sigma}(\zeta^i/R - \bar{\sigma}^i + \bar{\sigma}_k\, {}_1\lambda^k\, {}_1\lambda^i + \bar{\sigma}_k \nu_1{}^k \nu_1{}^i + \cdots + \bar{\sigma}_k \nu^k{}_{m-1} \nu^i{}_{m-1}),$$

where $\nu_1{}^i, \cdots, \nu^i{}_{m-1}$ are the $(m-1)$ normals to C in V_m.

If V_m contain a congruence of σ-curves, from (5.4) it is seen that $\zeta'^i = e^{-\sigma}\zeta^i$. Conversely, if the normal curvature vectors at points of corresponding curves C and C' are in corresponding directions, from the relation (2.6) written for C', we find after replacing primed quantities by means of unprimed that

$$(5.6) \qquad 1/R' = e^{-\sigma}(1/R - \bar{\sigma}_k \zeta^k).$$

Since V_m contains σ-curves, normal curvatures at points of C and C' are related by

$$1/R' = e^{-\sigma}(1/R)$$

and hence asymptotic lines in V_m and V'_m correspond. Conversely, if asymptotic lines C and C' correspond, it follows from (5.4) that the V_m contains a congruence of σ-curves. Accordingly: *A necessary and sufficient condition that asymptotic curves in V_m and V'_m correspond is that V_m contain a congruence of σ-curves.*

---

[6] E. G., p. 160.

If a curve C be a line of curvature for the normal ${}_{\tau}\xi^i$ it will satisfy the condition [7]

$$(5.7) \qquad {}_{\tau}\xi^i{}_{,j}\,{}_1\lambda^j = \rho\,{}_1\lambda^i.$$

From (1.5) and (5.7), it follows that the same condition holds for C' and the normal ${}_{\tau}\xi'^i$, i. e., *lines of curvature in V_m for a given normal correspond to lines of curvature in V'_m for the corresponding normal.*

If the normal curvature vector to a V_m (not containing σ-curves) at a point and in a given direction corresponds to the normal curvature vector at the corresponding point and in the corresponding direction, then, from (5.4), by means of (5.6), we have

$$\bar{\sigma}_k\zeta^k\zeta^i = \sum_{\tau=1}^{n-m}\bar{\sigma}_k\,{}_{\tau}\xi^k\,{}_{\tau}\xi^i.$$

Multiplying both sides by any one of the ${}^{\tau}\xi_i$ and summing on i, gives

$$(\bar{\sigma}_k\zeta^k\zeta^i - \bar{\sigma}^i)\,{}^{\tau}\xi_i = 0 \qquad (\tau = 1, \cdots, n-m).$$

This equation will be satisfied if the vector, $\bar{\eta}^i = \bar{\sigma}_k\zeta^k\zeta^i - \bar{\sigma}^i$, lies in V_m or if the direction σ^i coincide with the direction ζ^i. Conversely, if, for a given direction at a point, $\sigma^i = \zeta^i$ or if $\bar{\eta}^i$ lies in V_m, it follows from (5.5) that ζ^i for that direction will correspond to ζ'^i for the corresponding direction in V'_m. *The normal curvature vector to a V_m (not containing σ-curves) at a point and in a given direction will correspond to the normal curvature vector to V'_m at the corresponding point and in the corresponding direction, if and only if the vector $\bar{\eta}^i$ lie in V_m or the vector σ^i coincide with the vector ζ^i for the given direction.*

If normal curvature vectors correspond, it follows from (5.4) and (5.3) that

$${}^{\tau}\Omega_{\alpha\beta}\,{}_1\lambda^{\alpha}\,{}_1\lambda^{\beta}\,{}_{\tau}\xi^i = \rho\sum_{\tau=1}^{n-m}\bar{\sigma}_k\,{}_{\tau}\xi^k\,{}_{\tau}\xi^i$$

or, since the ${}_{\tau}\xi^i$ are linearly independent, that

$$({}^{\tau}\Omega_{\alpha\beta} - \rho\bar{\sigma}_k\,{}_{\tau}\xi^k a_{\alpha\beta})\,{}_1\lambda^{\alpha}\,{}_1\lambda^{\beta} = 0 \qquad (\tau = 1 \cdots n-m).$$

This equation says that *if the normal curvature vectors to V_m along C correspond to the normal curvature vectors along C'* FOR EVERY *C then V_m has completely indeterminate lines of curvature for all normals, ${}_{\tau}\xi^i$, and the normal curvatures in these $(n-m)$ directions, defined by*

---

[7] R. G., p. 168.

$$\frac{1}{R_\tau}=\frac{{}^\tau\Omega_{\alpha\beta}\,{}_1\lambda^\alpha\,{}_1\lambda^\beta}{a_{\alpha\beta}\,{}_1\lambda^\alpha\,{}_1\lambda_\beta},$$

are proportional to the cosines of the angles which the normals make with the σ-curves, or else V_m contains a congruence of σ-curves. The converse also holds.

The mean curvature of V_m for the normal ${}_\tau\xi^i$ is defined by[8]

$${}^\tau\Omega={}^\tau\Omega_{\alpha\beta}a^{\alpha\beta}$$

and is related to the mean curvature of V'_m for the normal ${}_\tau\xi'^i$ by

$$(5.8)\qquad {}^\tau\Omega'=e^{-\sigma}({}^\tau\Omega-\bar\sigma_k\,{}_\tau\xi^k).$$

The mean curvature normal of V_m is defined by

$$\bar\xi^i={}^\tau\Omega_{\alpha\beta}a^{\alpha\beta}\,{}_\tau\xi^i$$

and is related to the mean curvature normal of V'_m by

$$(5.9)\qquad \bar\xi'^i=e^{-2\sigma}(\bar\xi^i-{}_\tau\Sigma\bar\sigma_k\,{}_\tau\xi^k\,{}_\tau\xi^i).$$

The mean curvature of V_m is defined as the mean curvature of V_m for the mean curvature normal and is denoted by M. From (5.8), we have

$$(5.10)\qquad M'=e^{-\sigma}(M-\bar\sigma_k\xi^k).$$

From these relations we may state: *A necessary and sufficient condition that mean curvatures of V_m and V'_m be related by $M'=e^{-\sigma}M$, is that V_m contain a congruence of σ-curves.*

If equation (5.9) is treated as was (5.4), we find that mean curvature normals of V_m and V'_m correspond if the direction, $\bar\sigma_k\xi^k\xi^i-\bar\sigma^i$, lies in V_m or if σ^i coincide with the direction ξ^i at a point. From a relation similar to (5.5) obtained by replacing ζ^i by ξ^i, it is seen that the converse also holds. Hence: *The mean curvature normal to V_m (not containing σ-curves) corresponds to the mean curvature normal to V'_m if and only if the vector, $\bar\sigma_k\xi^k\xi^i-\bar\sigma^i$, lies in V_m or the direction σ^i coincide with the direction ξ^i.*

The principal normals of a curve in V_m and of the corresponding curve in V'_m are related by an equation similar to (2.4),

$$\nu'^i/\rho_g=e^{-2\sigma}(\nu^i/\rho_g+\bar\sigma_\gamma\,{}_1\lambda^\gamma\,{}_1\lambda^i-\bar\sigma^i),$$

where $\bar\sigma^i$ are the components in V_n of the direction $\bar\sigma^\alpha=a^{\alpha\beta}\bar\sigma_\beta=a^{\alpha\beta}(\partial\sigma/\partial y^\beta)$ in V_m. If equation (2.6) be written for C' and primed quantities be replaced by unprimed, we obtain

---

[8] R. G., p. 168.

$$\overline{\sigma}^i = \tilde{\sigma}^i + \sum_{\tau=1}^{n-m} \overline{\sigma}_k \, {}_\tau\xi^k \, {}_\tau\xi^i, \tag{5.11}$$

which says that the curves defined by the congruence $\tilde{\sigma}^i$, called $\tilde{\sigma}$-curves, are projections of the σ-curves on the V_m. These $\tilde{\sigma}$-curves play the same rôle in the subspace that the σ-curves do in V_n.

From the last theorem of section three and from the fact that the given conformal transformation induces a transformation with constant coefficient of magnification on a subspace if and only if that subspace lie in the H_σ, it follows that: *A necessary and sufficient condition that a geodesic C in V_m correspond to a geodesic in V'_m is that C be a σ-curve or that V_m lie in H_σ.*

If the V_m contain a congruence of σ-curves, a geodesic in V_m will correspond to a geodesic in V'_m if and only if it be a $\tilde{\sigma}$-curve. It is to be noted that if the subspace contain no σ-curves, no geodesic in both V_n and V_m can correspond to a geodesic in V'_n and V'_m.

6. Properties of curves in corresponding hypersurfaces. Inasmuch as there is but one normal ξ^i to a V_{n-1} in a V_n, and this normal corresponds in direction to the normal to V'_{n-1} in V'_n, the results of the previous section are somewhat simplified in this case. Theorems concerning asymptotic lines and lines of curvature may be paraphrased directly.

In considering the correspondence of geodesics, it is seen that if the V_{n-1} contain σ-curves, the only geodesics which can correspond to geodesics are the σ-curves. If the V_{n-1} be normal to the σ-curves, i. e., H_σ, all geodesics correspond to geodesics. For a general hypersurface, a necessary and sufficient condition that a geodesic correspond to a geodesic is that C be a $\tilde{\sigma}$-curve. Since principal normals to a geodesic in V_{n-1} are normal to V_{n-1}, it follows that a geodesic in V_{n-1} corresponds to a geodesic in V'_{n-1} if and only if C and C' have corresponding principal normals. It can be shown further that if C is a $\tilde{\sigma}$-curve (i. e., $\nu' = e^{-\sigma}\nu^i$, $\rho'_g/\rho_g = e^\sigma$) and if C and C' have corresponding principal normals, from the relation (2.6) written for C', by replacing the primed quantities in terms of the unprimed, then the principal normals of C are normal to the V_{n-1} and hence C and C' are geodesics in V_{n-1} and V'_{n-1}. Hence: *If a curve in a hypersurface V_{n-1} which does not contain a congruence of σ-curves and is not normal to the σ-curves, is a $\tilde{\sigma}$-curve, and if C and C' have corresponding principal normals, then C and C' are geodesic in V_{n-1} and V'_{n-1}, and conversely.*

7. Parallelism. From the relation (3.4), it appears that corresponding directions are simultaneously parallel along C and C' (not σ-curves) if and only if

$$\mu^i{}_{,j}\,{}_1\lambda^j = 0, \quad g_{ij}\sigma^i\mu^j = 0, \quad g_{ij}\,{}_1\lambda^i\mu^j = 0, \tag{7.1}$$

i. e., if and only if directions μ^i are parallel along C and are normal to the σ-surface of C.

If we write,

$$\begin{aligned} {}_1\lambda^i &= l^a\,{}_a\sigma^i \\ \mu^i &= m^a\,{}_a\sigma^i \end{aligned} \qquad (\alpha = 1 \cdots n), \tag{7.2}$$

where ${}_a\sigma^i$ is the normal ennuple of the σ-curves, then the conditions for simultaneous parallelism become

$${}_a\Sigma l^a m^a = 0, \tag{7.3a}$$

$$m^1 = 0, \tag{7.3b}$$

$$dm^a/ds = -\gamma^a{}_{\gamma\beta}m^\gamma l^\beta, \tag{7.3c}$$

where s is the arc of C and the γ's are the Ricci coefficients of rotation for the ennuple ${}_a\sigma^i$.

Differentiating (7. 3a) and $\gamma_{a\beta}m^\gamma l^\beta = 0$ with respect to s, and making use of (7. 3c), we obtain the following system of $(n+2)$ equations which the $2n$ quantities l^a, m^a must satisfy:

$$m^a(dl^a/ds) - \gamma^a{}_{\beta\gamma}l^a l^\gamma m^\beta = 0, \tag{7.4a}$$

$$(dl^\beta/ds)\gamma^1{}_{\gamma\beta}m^\gamma - \gamma^1{}_{a\beta}\gamma^\gamma{}_{\delta\epsilon}m^\delta l^\beta l^\epsilon + (\partial\gamma^1{}_{\gamma\beta}/\partial x^i)\,m^\gamma l^\beta\,{}_a\sigma^i l^a = 0, \tag{7.4b}$$

$$(dm^a/ds) + {}^{\gamma a}{}_{\gamma\beta}m^\gamma l^\beta = 0. \tag{7.4c}$$

If equations (7. 4a) and (7. 4b) are dependent, we find that

$$\gamma^1{}_{\gamma a} = 0 \qquad \gamma^1{}_{aa} = \rho \tag{7.5}$$

i. e., that ${}_1\sigma^i$ is a normal congruence, that ${}_2\sigma^i, \cdots, {}_n\sigma^i$ are canonical with respect to ${}_1\sigma^i$ and that lines of curvature of the hypersurfaces, H_σ, are completely indeterminate.[9] Conversely, if the hypersurfaces, H_σ, have completely indeterminate lines of curvature, the γ's for any orthogonal ennuple, and in particular for the normal ennuple of the σ-curves, satisfy the conditions

$$\gamma^1{}_{a\gamma} = 0, \qquad \gamma^1{}_{aa} = \gamma^1{}_{\beta\beta}$$

and a solution of the system (7. 4) will involve $(n-1)$ instead of $(n-2)$ arbitrary functions. Hence: *If the hypersurfaces, H_σ, have completely indeterminate lines of curvature, equations* (7. 4a) *and* (7. 4b) *are dependent and*

---

[9] R. G., p. 126

the general solution of equations (7.4) *involves* $(n-1)$ *arbitrary functions* $l^2, \cdots, l^n$; *otherwise a solution involves* $(n-2)$ *functions* $l^3, \cdots, l^n$. *In either case, the solution is uniquely determined by the arbitrary functions and a set of initial values satisfying the conditions,*

$$
\begin{aligned}
{}_0m^1 &= 0, & \gamma^1{}_{\gamma\beta}\, {}_0m^\gamma\, {}_0l^\beta &= 0,\\
{}_a\Sigma\, {}_0m^a\, {}_0m^a &= 1, & {}_a\Sigma\, {}_0m^a\, {}_0l^a &= 0.
\end{aligned}
$$

Every such solution l^a, m^a *of equations* (7.4) *determines a curve on a* σ*-surface, along which directions* μ^i *are parallel, and such that corresponding directions are parallel along* C'.

We shall now consider the geometric properties of curves (other than σ-curves) which admit simultaneous parallelism. We have already seen that the directions parallel along these curves must be normal to the σ-surfaces of the curves. For the normal ${}_\gamma\xi^i$ to a V_2 immersed in a V_n we have the relation [10]

$$
(7.6) \quad {}_\gamma\xi^i{}_{,j}\, {}_1\lambda^j = -{}^\gamma\Omega_{\alpha\beta}a^{\alpha\delta}x^i{}_{,\delta}\, {}_1\lambda^\beta + {}_\tau\Sigma\mu_{\tau\gamma/\beta}\, {}_\tau\xi^i\, {}_1\lambda^\beta \qquad \begin{matrix}(\alpha, \beta, \delta = 1, 2)\\ (\tau, \gamma = 1 \cdots n-m)\end{matrix}
$$

where [11]

$$
\mu_{\tau\gamma/\beta}\, {}_1\lambda^\beta = g_{ij}\, {}_\gamma\xi^j{}_{,k}\, {}_1\lambda^k\, {}_\tau\xi^i.
$$

If directions ${}_\gamma\xi^i$ are parallel along C, (7.6) reduces to

$$
(7.7) \qquad {}^\gamma\Omega_{\alpha\beta}a^{\alpha\delta}x^i{}_{,\delta}\, {}_1\lambda^\beta = 0.
$$

Multiplying this by σ_i and then by ${}^1\lambda_i$ gives respectively

$$
(7.8) \qquad
\begin{aligned}
{}^\gamma\Omega_{\alpha\beta}\, {}_1\lambda^\alpha\sigma^\beta &= 0,\\
{}^\gamma\Omega_{\alpha\beta}\, {}_1\lambda^\alpha\, {}_1\lambda^\beta &= 0.
\end{aligned}
$$

Since C is not a σ-curve, these equations will be satisfied only when C is a doubly counting asymptotic line for the normal ${}_\gamma\xi^i$. Hence we may state: *Necessary and sufficient conditions that directions,* ${}_\gamma\xi^i$, *along* C, *normal to its* σ*-surface, be parallel are:* 1) *that for the given normal,* C *be a doubly counting asymptotic line on its* σ*-surface, and* 2) *that the* $(n-3)$ *vectors* $\mu_{\tau\gamma/\beta}$ *determined by the given normal coincide with the directions of the normal to* C *in its* σ*-surface, i. e.,* $\mu_{\tau\gamma/\beta}\, {}_1\lambda^\beta = 0$, $(\tau = 1, \cdots, n-2;\ \tau \neq \gamma)$.

In three-space relation (7.6) reduces to

$$
(7.9) \qquad {}_\gamma\xi^i{}_{,j}\, {}_1\lambda^j = -{}^\gamma\Omega_{\alpha\beta}a^{\alpha\delta}x^i{}_{,\delta}\, {}_1\lambda^\beta.
$$

---

[10] R. G., p. 168.

[11] R. G., p. 160.

Hence, if $n = 3$, a necessary and sufficient condition that directions ${}_{\gamma}\xi^i$ normal to the σ-surface of C, be parallel along C, is that C be a doubly counting asymptotic line on its σ-surface.

If we differentiate covariantly the equations

$$g_{ij}\,{}_1\lambda^i\,{}_{\gamma}\xi^j = 0, \qquad g_{ij}\sigma^i\,{}_{\gamma}\xi^j = 0, \qquad\qquad (i, j = 1 \cdots n)$$

and assume that C is a doubly counting asymptotic line for the normal ${}_{\gamma}\xi^i$, i. e., that (7. 8) hold, it is found that ${}_2\lambda^i$ and $\sigma^i{}_{,k}\,{}_1\lambda^k$ lie in the σ-surface of C. If $n = 3$, and if ${}_2\lambda^i$ and $\sigma^i{}_{,k}\,{}_1\lambda^k$ lie in the σ-surface of C, it follows conversely that C is a doubly counting asymptotic line on its σ-surface. Accordingly: in V_3, a necessary and sufficient condition that a curve C (not geodesic in V_n), asymptotic on its σ-surface, be doubly counting asymptotic, is that the associate directions with respect to C of the σ-curves at points of C be directions in the σ-surface of C.

If C is geodesic in V_3 and if σ^i be parallel along C, C is a doubly counting asymptotic line on its σ-surface. Therefore: *In V_3, if the σ-curves admit among their transversals a geodesic, that geodesic is a doubly counting asymptotic line on its σ-surface.*

From these results we conclude that the only curves in V_3 which admit simultaneous parallelism are σ-curves, geodesics along which the σ-curves are parallel, and curves asymptotic on their σ-surfaces such that the associate directions of the σ-curves with respect to the C curves are directions in the σ-surfaces.

RANDOLPH-MACON WOMAN'S COLLEGE,
LYNCHBURG, VA.

SURFACES WHOSE ASYMPTOTIC CURVES ARE TWISTED CUBICS.*

By E. P. Lane and M. L. MacQueen.

1. Introduction. The purpose of this paper is to put on record some results relative to the problem of determining all analytic non-ruled surfaces whose asymptotic curves are twisted cubics. The problem is reduced to the integration of an ordinary differential equation. Some special cases are considered, in which interesting results can be deduced from this equation. Some examples of surfaces whose asymptotic curves are twisted cubics are discussed, and reference is made to Terracini's work on this subject.

2. Analytic Basis. This section summarizes portions of the classical analytical theory of the projective differential geometry of curves and surfaces which are used in later developments. In ordinary space, in which a point has projective homogeneous coördinates $x^{(1)}, \cdots, x^{(4)}$, the parametric vector equation of an analytic non-ruled surface is

$$x = x(u, v), \tag{1}$$

the parameters being u, v. If the asymptotic curves on the surface are the parametric curves, the coördinates x satisfy a system of two partial differential equations which can be reduced to the form

$$\begin{aligned} x_{uu} &= px + \theta_u x_u + \beta x_v, \\ x_{vv} &= qx + \gamma x_u + \theta_v x_v \end{aligned} \qquad (\theta = \log \beta\gamma), \tag{2}$$

subscripts indicating partial differentiation, and the coefficients being functions of u, v, which satisfy certain integrability conditions.

The parametric vector equation of an analytic curve is

$$x = x(t), \tag{3}$$

the parameter being t. These coördinates x satisfy an ordinary differential equation of the form

$$x^{iv} + 4p_1 x''' + 6p_2 x'' + 4p_3 x' + p_4 x = 0 \qquad (x' = dx/dt, \cdots), \tag{4}$$

the coefficients being functions of t. Let P_2, P_3, P_4 be defined by the formulas

---

* Received October 25, 1937.

$$
\begin{aligned}
&P_2 = p_2 - p_1^2 - p_1' \\
(5)\qquad &P_3 = p_3 - 3p_1p_2 + 2p_1^3 - p_1'', \\
&P_4 = p_4 - 4p_1p_3 + 6p_1^2p_2 - 6p_1'p_2 - 3p_1^4 + 6p_1^2p_1' + 3p_1'^2 - p_1'''.
\end{aligned}
$$

Then two invariants θ_3, θ_4 of the differential equation are defined by the formulas

$$
(6)\qquad
\begin{aligned}
\theta_3 &= P_3 - \tfrac{3}{2}P', \\
\theta_4 &= P_4 - \tfrac{81}{25}P_2^2 - 2P_3' + \tfrac{6}{5}P_2''.
\end{aligned}
$$

It is well known that the integral curves of equation (4) belong to linear complexes in case $\theta_3 = 0$, and are twisted cubics in case $\theta_3 = \theta_4 = 0$. Twisted cubics thus appear as a subclass of the class of all curves belonging to linear complexes.

3. Surfaces whose asymptotic curves belong to linear complexes. The problem of determining all non-ruled surfaces whose asymptotic curves belong to linear complexes has been completely solved. As our method of attack on the problem before us consists in selecting from these surfaces the class of surfaces whose asymptotic curves are twisted cubics, it will be useful to state here some known results relative to surfaces whose asymptotic curves belong to linear complexes.

It is known that, in case the integral surfaces of equations (2) have the property that their asymptotic curves belong to linear complexes, the coefficients β, γ in the equations can be specialized so that

$$
(7)\qquad \beta = \gamma = \frac{\sqrt{U'V'}}{U + V} \qquad (U'V' \neq 0),
$$

where U is an arbitrary function of u alone, and V of v alone. Moreover, the coefficients p, q are given by the formulas

$$
(8)\qquad
\begin{aligned}
2p &= 3l_{uu} - \tfrac{3}{2}l_u^2 - 3\beta l_v - U_1, \\
2q &= 3l_{vv} - \tfrac{3}{2}l_v^2 - 3\beta l_u - V_1,
\end{aligned}
$$

wherein l is defined by

$$
(9)\qquad l = \log \beta
$$

and U_1, V_1 by

$$
(10)\qquad U_1 = \frac{DU^2 + EU + F}{U'}, \qquad V_1 = \frac{DV^2 - EV + F}{V'},
$$

where D, E, F are arbitrary constants.

4. Conditions for twisted cubics. Analytic conditions necessary and sufficient that the asymptotic curves on a surface not only belong to linear complexes but actually are twisted cubics can now be computed. If the coefficients of equations (2) satisfy the conditions of Section 3, then the coefficients of the equation of the form (4) for the asymptotic u-curves can be calculated, and are found to be given by the formulas

$$\begin{aligned} 2p_1 &= -3l_u, \\ 6p_2 &= 11l_u^2 - 5l_{uu} - 2p - 3\beta l_v, \\ 4p_3 &= -2p_u + 6pl_u + 10l_u l_{uu} - 2l_{uuu} - 3\beta^3 - 6l_u^3 + 6\beta l_u l_v, \\ p_4 &= -p_{uu} + 4p_u l_u + pl_{uu} - \beta p_v - \beta^2 q - 3pl_u^2 + p^2 + 3\beta pl_v. \end{aligned} \tag{11}$$

Then the functions P_2, P_3, P_4 defined by the formulas (5) are found to be given by

$$\begin{aligned} 6P_2 &= \tfrac{1}{2}S + U_1, \\ 4P_3 &= \tfrac{1}{2}S' + U_1', \\ 4P_4 &= U_1^2 + 3U_1'' \qquad\qquad (U_1' = dU_1/du, \cdots), \end{aligned} \tag{12}$$

where S is the Schwarzian derivative of U with respect to u, defined by the formula

$$S = \left(\frac{U''}{U'}\right)' - \tfrac{1}{2}\left(\frac{U''}{U'}\right)^2. \tag{13}$$

The invariants θ_3, θ_4 defined by the formulas (6) can now be calculated for the u-curves. Of course we find $\theta_3 = 0$. But calculating θ_4 and setting the result equal to zero, we obtain the following *necessary and sufficient condition that the asymptotic u-curves be twisted cubics:*

$$S'' + \tfrac{3}{20}S^2 + \tfrac{3}{5}SU_1 - \tfrac{16}{15}U_1^2 - 3U_1'' = 0. \tag{14}$$

The analogous condition that the asymptotic v-curves be twisted cubics is

$$T'' + \tfrac{3}{20}T^2 + \tfrac{3}{5}TV_1 - \tfrac{16}{15}V_1^2 - 3V_1'' = 0, \tag{15}$$

where T is defined by

$$T = \left(\frac{V''}{V'}\right)' - \tfrac{1}{2}\left(\frac{V'}{V''}\right)^2. \tag{16}$$

Because of the analogy between equations (14) and (15) it will be sufficient to confine our discussion to equation (14). This is to be regarded as an ordinary differential equation of the fifth order for the determination of U as a function of u, and the problem of determining all non-ruled surfaces whose asymptotic curves are twisted cubics is thus in effect reduced to the solution of this differential equation.

5. Analytical theorems. A well-known theorem states that the general solution of the third-order differential equation $S=0$ is obtained by setting U equal to a linear fractional function of u with constant coefficients. In the course of our investigation, we have been led to some theorems of a kindred nature, which we have not found in the literature, and which we shall state here. The first three theorems relate to the Schwarzian derivative S defined by the formula (13), and their proofs, being immediate, will be omitted.

THEOREM 1. *The general solution of the differential equation*

$$S=k \qquad (k=\text{const.}\neq 0)$$

is obtained by integrating the differential equation

$$U'=PU^2+QU+R,$$

where P, Q, R are constants such that

$$Q^2-4PR=-2k.$$

THEOREM 2. *An integral of the differential equations*

$$S=\frac{AU^2+BU+C}{U'}=k \qquad (k=\text{const.}\neq 0)$$

is

$$U'=\frac{AU^2+BU+C}{k}$$

where A, B, C are constants such that $B^2-4AC=-2k^3$.

THEOREM 3. *An integral of the differential equation*

$$S=\frac{AU^2+BU+C}{U'}\neq\text{const.},$$

where A, B, C are arbitrary constants, is

$$\frac{AU^2+BU+C}{U'}\left(\frac{U''}{U'}\right)^2-2(2AU+B)\frac{U''}{U'} + 4AU'+\left(\frac{AU^2+BU+C}{U'}\right)^2=c' \tag{17}$$

where c' is an arbitrary constant.

It is also easy to verify the following statement:

THEOREM 4. *The Schwarzian derivative S defined by the formula* (13), *and the function U_1 defined by the first of the formulas* (10), *satisfy the equation*

$$U_1''' + 2SU_1' + S'U_1 = 0, \tag{18}$$

and also satisfy the equation

$$U_1'^2 - 2U_1U_1'' - 2U_1^2S = E^2 - 4DF. \tag{19}$$

6. Special cases. In certain special cases interesting conclusions can be drawn from equation (14), either alone or in the presence of the equations of Section 5. In particular, the following four theorems can be proved; the details of the demonstrations are so elementary as not to need to be reproduced here.

THEOREM 5. *If* $S = 0$, then $U_1 = 0$.

THEOREM 6. *If* $U_1 = 0$, *then* $S'^2 + \frac{1}{10}S^3 = \text{const.}$

THEOREM 7. *If* $S = k = \text{const.} \neq 0$, *then either* $U_1 = 3k/4$ *or* $U_1 = -3k/16$. *In the first alternative we have*

$$E^2 - 4DF = -\tfrac{9}{8}k^3, \qquad U' = \frac{DU^2 + EU + F}{3k/4},$$

and in the second,

$$E^2 - 4DF = -\tfrac{9}{128}k^3, \qquad U' = \frac{DU^2 + EU + F}{-3k/16}.$$

THEOREM 8. *If* $U_1 = \text{const.} \neq 0$, *then* $S = \text{const.} \neq 0$.

In the special case in which

$$S = T = U_1 = V_1 = 0,$$

we have $D = E = F = 0$, and U, V can be expressed in the form

$$U = \frac{au + b}{cu + d}, \qquad V = \frac{a'v + b'}{c'v + d'} \qquad (ad - bc \neq 0, a'd' - b'c' \neq 0),$$

in which $a, b, c, d, a', b', c', d'$ are arbitrary constants. Choosing a proportionality factor for these constants so that

$$ad - bc = 1, \qquad a'd' - b'c' = 1,$$

we are able to integrate equations (2) completely in this case and thus to prove the following theorem.

THEOREM 9. *The asymptotic curves are twisted cubics, and the directrix curves are indeterminate, on the surface whose parametric equations referred to its asymptotics are*

$$\begin{aligned} x_1 &= 1, \\ x_2 &= (\tfrac{1}{2}cu^2 + du) - (\tfrac{1}{2}c'v^2 + d'v), \\ x_3 &= (\tfrac{1}{2}au^2 + bu) + (\tfrac{1}{2}a'v^2 + b'v), \\ x_4 &= \tfrac{1}{6}(u^3 + v^3) + (\tfrac{1}{2}au^2 + bu)(\tfrac{1}{2}c'v^2 + d'v) \\ &\qquad + (\tfrac{1}{2}cu^2 + du)(\tfrac{1}{2}a'v^2 + b'v), \end{aligned}$$

and on every surface projectively equivalent to this one. These surfaces are algebraic and of order six or eight.

Another interesting special case is that characterized by the conditions

$$U_1 \neq \text{const.}, \quad S \neq \text{const.}, \quad U_1 = rS, \quad r = \text{const.} \neq 0. \tag{20}$$

In this case equation (18) gives at once

$$S''' + 3SS' = 0,$$

and integration then leads to

$$2S'' + 3S^2 = c', \tag{21}$$

where c' is an arbitrary constant. Substitution of our expressions for U_1 and S'' in equation (14) leads to an identity in S, from which we conclude that $c' = 0$ and that *the ratio r can have only one or the other of the two values $r = 9/2$ or $r = 9/32$.* Equation (19) yields

$$S'^2 + S^3 = \frac{E^2 - 4DF}{r^2}.$$

If we place

$$A = D/r, \qquad B = E/r, \qquad C = F/r,$$

and refer to Theorem 3 we see that equation (17) with $c' = 0$ is valid.

7. Examples. Three examples of surfaces whose asymptotic curves are twisted cubics and which have been considered by different geometers, will now be adduced. It happens that these examples belong to special cases mentioned above.

The asymptotic curves on the minimal surface of Enneper are known[1] to be twisted cubics. Parametric equations of this algebraic surface of order nine, referred to its asymptotic curves, are

$$\begin{aligned} x_1 &= 1, \\ x_2 &= (u + v)[3 + 3(u - v)^2 - (u + v)^2], \\ x_3 &= (u - v)[3 + 3(u + v)^2 - (u - v)^2], \\ x_4 &= 12uv. \end{aligned} \tag{22}$$

---

[1] Darboux, *Leçons sur la théorie générale des surfaces*, second edition, vol. I (1914), pp. 374-376.

Application of our general theory to this surface offers no difficulty. Omitting the calculations, we shall merely state that this surface belongs to the special case characterized by the conditions (20) with $r = 9/32$.

Wilczynski remarked [2] that the surface whose parametric equations, referred to its asymptotics, are

$$
\begin{aligned}
&x_1 = 1, \qquad x_2 = u + v, \qquad x_3 = U' - V', \\
&x_4 = -2(U' - V')(u - v) + 4(U + V),
\end{aligned} \tag{23}
$$

where U and V are cubic polynomials in u alone and v alone, respectively, is an algebraic surface of the sixth order on which the directrix curves are indeterminate and on which the asymptotic curves are twisted cubics. Application of our general theory to this surface of Wilczynski shows that it belongs to the special case characterized by the conditions $S = U_1 = 0$, so that U is a linear fractional function of u, and $D = E = F = 0$.

Enriques has studied [3] the surface whose algebraic equation is

$$
(x_1 x_3 - x_2^2)^3 k = (x_1^2 x_4 + 2x_2^3 - 3x_1 x_2 x_3)^2 \qquad (k = \text{const.}), \tag{24}
$$

which has also been investigated [4] recently by Emma Castelnuovo. As Enriques pointed out, this surface has not only the property that it admits a two-parameter family of projective transformations into itself, but also the property that its asymptotic curves are twisted cubics. Parametric equations of this surface of Enriques, referred to its asymptotic curves, are

$$
\begin{aligned}
&x_1 = 1, \quad x_2 = \frac{1+h}{2} u + \frac{1-h}{2} v, \quad x_3 = \frac{1+h}{2} u^2 + \frac{1-h}{2} v^2, \\
&x_4 = \frac{(1+h)^2}{4h} u^3 - 3 \frac{1-h^2}{4h} u^2 v + 3 \frac{1-h^2}{4h} u v^2 - \frac{(1-h)^2}{4h} v^3,
\end{aligned} \tag{25}
$$

where

$$
h^2 = \frac{4}{4-k}.
$$

Application of our general theory shows that this surface is a special surface of Wilczynski. In fact, the function U of the general theory is linear in u for the surfaces of Enriques, whereas U is a linear fractional function of u for the general surface of Wilczynski.

---

[2] Wilczynski, Abstract, *Bulletin of the American Mathematical Society*, vol. 20 (1913-14), p. 312.

[3] Enriques, "Le superficie con infinite trasformazioni proiettive in sè stesse," *Atti del R. Istituto Veneto di scienze, lettere ed arti*, ser. 7, vol. 4 (1893), pp. 1590-1635. See also Enriques, Intorno alla Memoria "Le superficie con infinite trasformazioni proiettive in sè stesse," *ibid.*, vol. 5 (1894), and Lie, "Bestimmung aller Flächen, die eine continuerliche Schaar von projectiven Transformationen gestatten," *Berichte der Gesellschaft der Wissenschaften zu Leipzig*, vol. 47 (1895).

[4] E. Castelnuovo, "Di una classe di superficie razionali che ammettono ∞^2 trasformazioni proiettive in sè," *Rendiconti dei Lincei*, ser. 6, vol. 24 (1936), pp. 342-346.

8. Terracini's formulas. Terracini has shown [5] that if a non-ruled surface has the property that its asymptotic curves belong to linear complexes, then parametric equations of the surface, referred to its asymptotics, can be written in one or another of the following three forms, according to the nature of a certain quadric, commonly called the quadric of Sullivan, associated with the surface:

$$
(26)\qquad
\begin{array}{ll}
x_1 = (V' - U')u + 2U, & x_2 = (V' - U')v - 2V, \\
x_3 = (V' - U')uv - 2(Vu - Uv), & x_4 = V' - U'; \\[1ex]
x_1 = (U' - V')u - 2U, & x_2 = (U' - V')v + 2V, \\
x_3 = 2uv, & x_4 = -2; \\[1ex]
x_1 = (U' - V')(u - v) - 2(U + V), & x_2 = U' - V', \\
x_3 = u + v, & x_4 = 1,
\end{array}
$$

where U is an arbitrary function of u alone, and V of v alone. In all three cases the coefficients of the differential equation of the form (4) for the u-curves are found by actual calculation to be given by the formulas

$$(27)\qquad 2p_1 = -R, \qquad 6p_2 = -R' + R^2, \qquad p_3 = p_4 = 0,$$

where R is defined by placing

$$(28)\qquad R = \frac{U^{iv}}{U'''}.$$

Calculating the invariant θ_4 for this equation and setting this invariant equal to zero, we obtain the following differential equation:

$$(29)\qquad 10R''' - 30RR'' + 32R^2R' - 19R'^2 - 4R^4 = 0.$$

There is of course a similar equation for the v-curves. Thus the problem of determining parametric equations of all non-ruled surfaces whose asymptotic curves are twisted cubics is reduced to the problem of integrating the differential equation (29) of the third order to obtain the function R, and then performing four quadratures on equation (28) to obtain the function U.

THE UNIVERSITY OF CHICAGO.

---

[5] Terracini, "Sulle superficie le cui asintotiche dei due sistemi sono cubiche sghembe," *Atti della Società dei Naturalisti e Matematici di Modena*, ser. 5, vol. 5 (1919-20).

ON HYPERGROUPS, MULTIGROUPS, AND PRODUCT SYSTEMS.*

By L. W. Griffiths.

1. Introduction.

In this paper there is considered an abstract system whose elements are classes. Each class is an unordered set of marks, selected from a fundamental set Σ of distinct marks. The marks in a class are distinct if and only if this property is specifically stated for the system. The classes are distinct.

A product system is a system satisfying two postulates. First, the system is closed with respect to an addition process which is associative and commutative. Second, the system is closed with respect to a multiplication process which is associative, and which is distributive, on the right and on the left, with respect to the addition process. A division system Δ is a product system which satisfies postulate III or postulate III'. Thus, in particular, if a product system contains the subset Ω of all classes each of which consists of exactly one mark, then for every pair, A and B, of classes in Ω there are two classes, B' and B'', each in Ω, such that BB' contains A and $B''B$ contains A.

A Γ system is formed from a division system which contains Ω by replacing each class in Ω by the mark of which that class consists. It is proved that the fundamental set Σ in a Γ system is a hypergroup as defined by Marty.[1] Conversely, a Marty hypergroup is embedded as Σ in the Γ system formed from its product system. If it is postulated that a Γ system has a mark in Σ whose properties are suggested by those of the identity in group theory, then Σ is a regular hypergroup. If it is postulated that this mark is unique, and that for each mark in Σ there is in Σ a mark whose properties are suggested by those of inverse elements in group theory, then Σ is a completely regular hypergroup. A further condition makes Σ a normal hypergroup of Marty.[2] Conversely, each of these types of hypergroups is embedded in the Γ system of its products, and this Γ system has the corresponding property stated above.

It is proved that a multigroup as defined by Ore [3] is a completely regular

---

* Received November 17, 1937; Revised slightly, December 20, 1937.

[1] F. Marty, "Sur une généralisation de la notion de groupe," *Särtryck ur Förhandlinger vid Åttonde Skandinaviska Matematikerkongressen i Stockholm* 1934, pp. 45-49.

[2] F. Marty, "Rôle de la notion d'hypergroupe dans l'étude des groupes non abeliens," *Comptes Rendus de l'Academie des Sciences*, vol. 201 (1935), pp. 636-638.

[3] O. Ore, "Structures and group theory I," *Duke Mathematical Journal*, vol. 3 (1937), pp. 149-174.

hypergroup of Marty for which multiplication by the identity on one side is unique.

If a Γ system satisfies the preceding conditions for a Marty completely regular hypergroup except that there may be more than one identity element in Σ, and if there exists a positive integer n such that the product of every pair of marks in Σ is a class consisting of precisely n marks, then Σ is a hypergroup as defined by Wall.[4] Conversely, a Wall hypergroup is a special kind of Marty hypergroup, namely one for which such an integer n exists and for which an identity and inverse elements exist, and hence is embedded in a Γ system as is a Marty hypergroup.

A division system which contains Ω particularizes to a group if it is merely Ω and if in postulate III B' and B'' are unique. Then the embedded hypergroup is the group. A division system which does not contain Ω does not particularize to a group whose elements are marks in Σ. However, it is proved in Theorem 7 that if the classes of such a division system are regarded as marks of a new fundamental system Σ', then the product system of Σ' is Σ'. Hence either a division system contains a Marty hypergroup Ω or this division system as Σ' is a Marty hypergroup in which multiplication is unique. Thus either a division system particularizes to a group whose elements are the marks in Σ or it particularizes to a group whose elements are the marks in Σ'.

In section 7 it is proved that the direct product of two finite groups is simply isomorphic with a subset of classes in a division system which does not contain Ω. However, as stated above, this subset of classes is a subgroup of product system Σ'.

A division system is a generalization of the abstract system group, since it has the closure property for multiplication, since either division is postulated or the existence of elements analogous to the identity and inverse elements of group theory is postulated, and since it particularizes to a group in that one of the two senses explained above which corresponds to that one of postulates III or III′ used in its definition. Furthermore, as proved in section 3, if further conditions are to be placed on a product system, of a type suggested by the usual postulates for a group, a division system is the most general type so obtained.

There is defined a Σ algebra, and therefore in particular a hypergroup algebra, which is analogous to the ordinary group algebra.

2. Postulational definition of a Δ system and of a Γ system. Consider a set Σ of distinct marks, $a, b, c, \cdots$. If k is a positive integer, then $a_1, \cdots, a_k$

---

[4] H. S. Wall, "Hypergroups," *American Journal of Mathematics*, vol. 59 (1937), pp. 77-98.

is a notation for a finite set of k of these marks, and $a_1, a_2, \cdots$ is a notation for an infinite set. In these sets the marks are not necessarily distinct. R, S, T, M, and other capital letters near the end of the alphabet represent classes, $\{a_1, \cdots, a_k\}$ and $\{a_1, a_2, \cdots\}$. The marks in a class are not ordered, nor are they necessarily distinct. A, B, C, A_1, and other capital letters near the beginning of the alphabet, with or without subscripts, represent classes each of which consists of exactly one mark; thus, for example, A is the class $\{a\}$, A_1 the class $\{a_1\}$. Ω is the set of all such classes A. Thus Ω is in one to one correspondence with Σ. The class R may be a class in Ω, in particular. A system of classes may or may not contain classes in Ω.

Two classes are equal if and only if the set of marks in the one class is the set of marks in the other. The class R includes the class S, in notation $R \supset S$, means that the set of marks in S is a subset of the set of marks in R. $R \supset S$ does not exclude $R = S$.

An addition of classes which is illustrated in sections 4 and 5 of this paper is the following. The system of classes allows no repetition of marks in a class. The sum of two classes is the class whose set of marks is the set of distinct marks among the set of all the marks of the first class and all those of the second class. An addition of classes for a system in which repetition of marks in a class is allowed is illustrated in section 6. It is the following: the sum of two classes is the class whose set of marks is the set of all the marks of the first class and all those of the second class. Each of these addition processes is associative and commutative.

A product system is a system of distinct classes which satisfies the first two of the following postulates.

Postulate I. For every pair, R and S, of classes in the system there is in the system a unique class $R + S$. This addition process is associative and commutative.

Postulate II. For every pair, R and S, of classes in the system there is in the system a unique class RS. This multiplication process is associative, and it is distributive, on the right and on the left, with respect to the addition process of postulate I.

Postulate III. The system contains Ω, and for every pair, A and B, of classes in Ω there are two classes, B' and B'', each in Ω, such that $BB' \supset A$ and $B''B \supset A$. B' is not necessarily unique, nor is B''.

Postulate III'. The system does not contain Ω (although it may contain one or more classes in Ω), and for every pair, R and S, of classes in the system there are two classes S' and S'', each in the system, such that $SS' \supset R$ and $S''S \supset R$. S' is not necessarily unique, nor is S''.

A division system Δ is a product system which satisfies postulates I, II, and III, or postulates I, II, and III′. A Γ system is formed from a Δ system which contains Ω in the following manner. From the definition of A in Ω a one to one correspondence of Ω to Σ is established by $A \leftrightarrow a$ if and only if $A = \{a\}$. A Γ system is formed from a Δ system which contains Ω by replacing each class A in Ω by the corresponding mark in Σ. For a Γ system addition and multiplication are defined, and postulates I, II, and III stated, by making these replacements in the definitions and postulates for the Δ system. A Γ system is simply isomorphic to the Δ system from which it is derived, and the correspondence is preserved under multiplication and addition. To Ω in the Δ system corresponds Σ in the Γ system. A corresponding Δ system and Γ system have precisely the same properties.

3. Related systems. Certain systems are considered having classes whose properties are suggested by those of the identity and inverse elements in group theory. They are related to Δ systems in the following theorems.

THEOREM 1. *If a system of distinct classes contains Ω, and satisfies postulates I and II, and the following postulates IV and V, then it satisfies postulate III.*

Postulate IV. There exists a class E in Ω such that for every A in Ω it is true that $AE \supset A$ and $EA \supset A$. (E is not necessarily unique.)

Postulate V. There exists a class E satisfying postulate IV such that for every A in Ω there exist A_1 and A_2 in Ω such that $AA_1 \supset E$ and $A_2A \supset E$. (A_1 is not necessarily unique, nor is A_2).

To prove Theorem 1, let A and B be arbitrary classes in Ω. Then by the associative and distributive laws applied to the equations corresponding to the inclusions it is seen that $BB_1A \supset EA \supset A$. Therefore there is a class B', such that $B' \subset B_1A$ and B' is in Ω, and that $BB' \supset A$. Similarly there is a class B'' such that B'' is in Ω and $B''B \supset A$. Hence postulate III is satisfied.

THEOREM 2. *If a system of distinct classes does not contain Ω, and if it satisfies postulates I and II and the following postulates VI and VII, then it satisfies postulate III′.*

Postulate VI. There exists a class M in the system such that for every class R in the system it is true that $RM \supset R$ and $MR \supset R$. (M is not necessarily unique.)

Postulate VII. There exists a class M satisfying postulate VI such that for every R in the system there exist two classes R_1 and R_2, each in the system, such that $RR_1 \supset M$ and $R_2R \supset M$. (R_1 is not necessarily unique, nor is R_2.)

To prove Theorem 2, let R and S be arbitrary classes in the system. Then if S_1 and S_2 are determined in accordance with postulate VII, by the associative and distributive laws and postulate VI, it is true that S_1R and RS_2 have the same properties as S' and S'' respectively of postulate III′.

THEOREM 3. *If a system of distinct classes contains* Ω, *and satisfies postulates I, II, VI and VII, then it satisfies the following postulate* III_0, *and then it satisfies postulate III.*

Postulate III_0. For every pair, R and S, of classes in the system there are two classes, S' and S'', each in the system, such that $SS' \supset R$ and $S''S \supset R$. (S' is not necessarily unique, nor is S''.)

To prove Theorem 3, let A and B be arbitrary classes in Ω. Then III_0 holds as in the proof of Theorem 2. Then by postulate III_0 there exists a class X in the system such that $BX \supset A$. Then by the distributive law there is a class B' in Ω such that $BB' \supset A$. Similarly there is a class B'' in Ω such that $B''B \supset A$.

THEOREM 4. *If a system of classes satisfies postulates II and III, and the following postulate I′, then it satisfies* III_0.

Postulate I′. The system is closed with respect to an addition process which is associative and commutative, and if it contains any class containing infinitely many marks then it contains all classes containing infinitely many marks.

To prove Theorem 4, let R and S be arbitrary classes in the system. If $R = A_1 + \cdots + A_k$, and $S = B_1 + \cdots + B_j$, and if A_r and B_s are arbitrary summands in R and S respectively, then there exist by postulate III classes C_{rs} in Ω such that $B_sC_{rs} \supset A_r$. Hence the class whose summands are precisely all these classes C_{rs} is effective as S' in postulate III_0. Similarly there exists a class S'', if both R and S are finite classes. If either or both of R and S are infinite classes, postulate I is insufficient to guarantee such a sum of classes C in the system, but the last part of postulate I′ does insure the sum in the system.

THEOREM 5. *If a system of distinct classes satisfies postulates I′, II, IV and V and contains* Ω*, then it satisfies postulates VI and VII.*

The proof of Theorem 5 is similar to the proofs of the preceding theorems. Postulate I was insufficient for infinite classes R, while postulate I′ was sufficient.

Therefore, if a system contains Ω and further conditions are to be placed on it, of a type suggested by the usual postulates for a group, a more general system is obtained under postulates I, II, and III than under various combinations from I or I′, II, III_0, or IV and V, or VI and VII.

The following theorem is useful later in relating the properties of these systems to those of a Marty hypergroup, to those of a Wall hypergroup, and to those of a multigroup.

THEOREM 6. *If a system of distinct classes contains* Ω *and if the addition process satisfying postulate I is either of the two addition processes defined preceding postulate I, then a multiplication process satisfies postulate II if and only if it satisfies the following postulate II′.*

Postulate II′. For every A and B in Ω there is a unique class AB in the system. If S and T are two classes in the system, one of which at least is not in Ω, then ST is the sum of all products AB as A ranges over all summands A in S and B ranges over all summands B in T. For every A, B, and C in Ω it is true that $(AB)C = A(BC)$.

To prove Theorem 6, let R and S be arbitrary classes in the system. Then by postulate II there is a unique class RS in the system. Hence, in particular, the first statement in postulate II′ is true. The proof that the second statement in postulate II′ is true should be read first under the definition of addition for the case that no repetition of summands is allowed, and then under the definition of addition for the case that repetition is allowed. Let $S = A_1 + \cdots$ and $T = B_1 + \cdots$ be two arbitrary classes, at least one of which, for example T, is not in Ω. Then by the distributive law $ST = SB_1 + \cdots$. If S is in Ω then this is precisely the second statement in postulate II′. If S is not in Ω, then again by the distributive law $ST = (A_1B_1 + \cdots) + (A_2B_1 + \cdots) + \cdots$; this is precisely the second statement in postulate II′. The converse will be proved before any consideration of associativity of multiplication is taken. Let S, T, and U have summands A_i, B_j and C_k respectively. Then by the first two statements in postulate II′ it is true that $ST = \Sigma A_iB_j$ summed for all pairs with A_i in S and B_j in T; $SU = \Sigma A_iC_k$ summed for all pairs A_i in S and C_k in U; $S(T + U) = \Sigma A_iD_k$ summed for all pairs A_i in S and D_k in $T + U$. Then clearly $ST + SU = S(T + U)$.

To prove that postulate II implies the third statement in postulate II′ it is to be noted that, since multiplication is associative for all classes under postulate II, it is so in particular for all classes in Ω. Conversely, if postulate II′ holds then it is proved as follows that multiplication for all classes is associative. For, by the preceding, multiplication of classes is distributive with respect to addition. Then, with the notations already introduced for S, T, and U, it is true that $(ST)U = [\Sigma(A_iB_j)]U = \Sigma[(A_iB_j)C_k]$ where the first summation is with respect to all pairs A_i in S and B_j in T and the second summation is with respect to all triples A_i in S, B_j in T, and C_k in U. Similarly $S(TU) = \Sigma[A_i(B_jC_k)]$ where the summation is with respect to these same triples. Since the class $A_i(B_jC_k)$ is the class $(A_iB_j)C_k$ by postulate II′, it follows that $(ST)U = S(TU)$. This completes the proof of Theorem 6.

4. Relation of a Γ system to a Marty hypergroup. Consider a Γ system derived from a Δ system satisfying postulates I, II, and III and allowing no repetition of marks in a class. Then Σ is a hypergroup as defined by Marty. For the letters $A, B, C, \cdots$ in Marty's notation mean the marks $a, b, c, \cdots$ in this paper. He explicitly states postulate III, and the first and last sentences of postulate II′. No explicit definition is given of the expression $(AB)C$, for example, nor is it explained why from $AE \supset A$ it follows that $AEB \supset AB$. But an analysis of the proofs indicates that the matter can be treated by introducing classes and class addition, and by defining the product of two classes as in the second statement of postulate II′. No explicit statement is made by Marty that no repetition of marks is allowed in products. However this is implicit, for the finite case at least, in the proof that $\sum_{i=1}^{i=n} \alpha_i{}^A = n$, since it is not permitted that i be greater than n there. Hence, by Theorem 6, the Σ of a Γ system derived as stated in the first statement of this section is a hypergroup, and conversely a Marty hypergroup is embedded in the Γ system of its own product classes.

It follows immediately that, if a system satisfies the hypotheses of Theorem 1 with E unique and $A_1 = A_2$, the set Σ in the derived Γ system is a Marty completely regular hypergroup; and that a completely regular hypergroup is embedded in such a system of its product classes. If $EA = A$ and $AE = A$, then Σ is a Marty normal hypergroup.

5. Relation of a Γ system to an Ore multigroup. The letters $B_1, B_2, \cdots$ in Ore's notation mean the marks $b_1, b_2, \cdots$ in this paper. Repetition of marks in a product is not allowed, since the marks in a product constitute a subset of the elements of the multigroup, that is, of Σ. In the example of a multigroup given by the co-sets of a group with respect to an arbitrary subgroup addition is used, but addition is not explicitly used in the general definition of a multigroup. Postulate II′ is explicitly stated, with the word "sum" replaced by "subset." Postulate IV is stated, with E unique and one of the inclusions an equality; postulate V is stated, with $A_1 = A_2$. Hence by Theorems 6 and 1 an Ore multigroup is a completely regular hypergroup of Marty for which multiplication by the identity on one side is unique. Hence if a Γ system satisfies postulates IV and V with these extra conditions on E and A_1, then Σ is a multigroup, and, conversely, a multigroup is embedded in such a Γ system of its products.

6. Relation of a Γ system to a Wall hypergroup. The letters $a, b, c, \cdots$ in Wall's notation mean the marks $a, b, c, \cdots$ in this paper. Repetition of marks is allowed in products. Addition is explicitly used, although no formal postulate is stated. Postulates II′, IV, and V with $A_1 = A_2$ are stated. There is introduced a positive integer n such that, in my notation, every product AB is a sum of exactly n classes in Ω. The bracket product is the class of all distinct marks in the ordinary product. Hence a Wall hypergroup of dimension n with bracket product multiplication is a Marty hypergroup with the condition of dimensionality imposed, and as such has the relation to a Γ system stated in section 4. A Wall hypergroup of dimension n with ordinary product multiplication is embedded as a set Σ in a Γ system of its own product classes. An nG_1 is a Wall hypergroup for which E is unique, and for which $A_1 = A_2$ and is unique for every A. The scalars, if existing, are a subset in a Wall hypergroup, and hence a subset in the Σ of the Γ system of the product classes.

7. The Γ system of a Δ system. The direct product of two finite groups embedded in a Δ system. The classes which constitute a Δ system may be regarded as a new set Σ' of marks. Since the product of two classes in a Δ system is a unique class in the Δ system, the product of two marks in Σ' is a unique mark in Σ', and the system of product classes is merely the set Ω' of all classes each of which consists of exactly one mark of Σ'. Hence postulates I and II hold for Ω'. If the Δ system does not contain Ω, then by postulate III′ for the Δ system it is true that postulate III holds for Ω'. Hence the Γ system of the Δ system Σ' is precisely Σ', that is, a Marty hypergroup in which multiplication is unique. This is a group, as Marty proved, if and only if in postulate III′ for the Δ system it is true that every S' is unique or every S'' is unique. Hence the first part of Theorem 7 is proved.

THEOREM 7. *If a Δ system does not contain its Marty hypergroup Ω, then the Γ system of this Δ system is this Δ system, and hence a Marty hypergroup in which multiplication is unique. If a Δ system contains Ω but is not equal to Ω, then the Γ system of this Δ system is this Δ system if and only if postulate III_0 holds for this Δ system. If a Δ system is Ω, then it is a Marty hypergroup in which multiplication is unique.*

The second statement in Theorem 7 is true, since postulate III holds for Σ' if and only if postulate III_0 holds for the Δ system.

The direct product of two finite groups, G with elements $a_1, \cdots, a_m$ and H with elements $b_1, \cdots, b_n$, is simply isomorphic with a set of classes embedded in a Δ system. Thus let Σ be $a_1, \cdots, a_m, b_1, \cdots, b_n$; it is no limitation to assume that a_1 is the identity of G and b_1 the identity of H, and that as marks a_1 and b_1 are distinct, although G and H might be subgroups in another group as elements of which a_1 and b_1 are the same. Consider the set of all classes each of which contains at least one mark from among $a_1, \cdots, a_m$ and at least one mark from among $b_1, \cdots, b_n$, with no repetition of marks allowed. This set of classes satisfies postulate I with addition defined as stated for the case that no repetition of marks is allowed. The number of these classes is finite. If R and S are two of these classes then RS is defined as the class whose marks are the distinct ones among all products a_ia_j and b_rb_s as a_i ranges over all the marks among $a_1, \cdots, a_m$ appearing in R, a_j ranges similarly over those in S, b_r ranges over all the marks among $b_1, \cdots, b_n$ appearing in R, and b_s ranges similarly over those in S. Postulate II is satisfied. In fact multiplication is commutative. This system of classes does not contain Ω. It is seen as follows that postulate III′ is satisfied. Let R and S be arbitrary classes. Let a and b be two arbitrary, fixed marks in S. For each mark a_i in S there is a unique mark a_j such that $aa_j = a_i$; for each mark b_r in S there is a unique mark b_s such that $bb_s = b_r$. Define S' as the set whose marks are these a_j and these b_s as a_i and b_r range over R. Then S' is such that $SS' \supset R$. It does not follow that S' is unique. Similarly a class S'' can be found.

Hence this set of distinct classes is a finite Δ system which does not contain Ω. The subset of all classes $\{a, b\}$ is simply isomorphic to the direct product of G and H. By Theorem 7 the Γ system of this Δ system is this Δ system; this Γ system is a Marty hypergroup containing the direct product group.

8. Definition of a Σ algebra; a hypergroup algebra. Consider a set of classes satisfying postulates I and II, but not necessarily postulate III or III′, that is, a set of classes which is a product system but not necessarily a Δ system. Let the product system contain Σ, by the process described at the end of section 2. Hence every class in the system is a sum of marks in Σ. Hence if, for every class R in the system (and hence in particular for every mark in Σ) and for every positive integer k, the notation kR means $R + \cdots + R$ in which R is a summand k times, and if $Rk = kR$, then every class in the system is a sum, in one and only one way, with coefficients which are positive integers, of marks in Σ. In particular, the product of two marks in Σ is such a sum. Hence the marks in Σ can be used as the basis of an algebra over an arbitrary field, for which the multiplication table of the basis elements is the multiplication table of the marks in Σ in the product system. This algebra may be called a Σ algebra. If the product system satisfies postulate III, then by section 4 the Σ algebra is an algebra whose basis elements form a hypergroup. Conversely, if an algebra is defined with the elements of a hypergroup as basis elements, then this algebra is such a Σ algebra. Such a Σ algebra may be called a hypergroup algebra. If the product system is in fact merely a group then the Σ algebra becomes a group algebra, since Σ is in fact a group.

The properties of elements of algebras are therefore in particular properties of marks in Σ, and hence in particular of elements of a hypergroup. For example, the matrix representation of hypergroups, as presented in section 5 of the paper by Wall, is a very special instance of a standard elementary property of algebras.[5]

NORTHWESTERN UNIVERSITY.

---

[5] M. Deuring, *Algebren*, Berlin 1935, p. 2; or B. L. v. d. Waerden, *Moderne Algebra*, Berlin 1931, vol. II, p. 131.

MATRICES NORMAL WITH RESPECT TO AN HERMITIAN MATRIX.*

By JOHN WILLIAMSON.

Introduction. A square matrix A with elements in the complex number field is said to be a normal matrix, if it is commutative with its conjugate transposed matrix, i. e., if

$$AA^* = A^*A,$$

where $A^* = \bar{A}'$ is the conjugate transposed of A. In particular every hermitian matrix and every unitary matrix is a normal matrix. A necessary and sufficient condition, that a matrix A be a normal matrix, is that A be equivalent under a unitary transformation to a diagonal matrix. In other words a matrix A is normal, if, and only if, there exists a unitary matrix V, such that

$$VAV^* = D, \tag{1}$$

where D is a diagonal matrix, i. e. a matrix all of whose elements not in the leading diagonal are zero.[1]

If $d_1, d_2, \cdots, d_m$ are the distinct latent roots or characteristic numbers of D and ϕ_i is the principal idempotent element of D associated with d_i, $i = 1, 2, \cdots, m$,[2]

$$D = d_1\phi_1 + d_2\phi_2 + \cdots d_m\phi_m.$$

Since ϕ_i is a diagonal matrix, all of whose elements are zero or unity, $\phi^*_i = \phi_i$ and consequently

$$D^* = \bar{d}_1\phi_1 + \bar{d}_2\phi_2 + \cdots \bar{d}_m\phi_m. \tag{2}$$

As ϕ_i is a polynomial in the matrix D, so that $\phi_i = \phi_i(D)$, it is a consequence of (2) that

$$D^* = \bar{d}_1\phi_1(D) + \bar{d}_2\phi_2(D) + \cdots + \bar{d}_m\phi_m(D) = g(D), \tag{3}$$

and of (1) and (3), that

$$A^* = g(A), \tag{4}$$

---

* Received February 1, 1938.

[1] Aurel Wintner, *Spektraltheorie der unendlichen Matrizen* (1929), p. 24.

[2] J. M. Wedderburn, "Lectures on matrices." *Colloquium Publications* (1934), p. 29.

where $g(x)$ is a polynomial in x. Hence, if A is a normal matrix, the conjugate transposed of A is a polynomial in A. Conversely, let $A^* = g(A)$. Then we may so determine the unitary matrix V in (1) that D is a triangle matrix, in which all of the elements below the leading diagonal are zero.[3] Since V is a unitary matrix, $D^* = g(D)$ and consequently each element of D^*, which lies below the leading diagonal, is zero. As D^* contains no elements, which are different from zero, above the leading diagonal, D is a diagonal matrix. Hence A is a normal matrix. We have therefore proved,

LEMMA 1. *A necessary and sufficient condition that a matrix A be a normal matrix is that $A^* = g(A)$, where $g(x)$ is a polynomial in x.*

By taking the conjugate transposed of the matrices in (4) we see that

$$A = \bar{g}(A^*) = f(A^*), \tag{5}$$

where, if $g(x) = \sum_{i=0}^{n} a_i x^i$, $f(x) = \bar{g}(x) = \sum_{i=0}^{n} \bar{a}_i x^i$. We have therefore, on combining (4) and (5),

$$A = f\{\bar{f}(A)\} = \bar{g}\{g(A)\}. \tag{6}$$

In particular, if A is hermitian, $f(x) \equiv x$ and, if A is unitary, $f(x) \equiv \phi(x)$, where $\phi(A) = A^{-1}$.

Let A be a normal matrix, so that A satisfies (5) and (6). If P is any non-singular matrix and if

$$B = PAP^{-1},$$

then

$$A = P^{-1}BP \quad \text{and} \quad A^* = P^*B^*(P^{-1})^*.$$

Therefore, as a consequence of (5),

$$P^{-1}BP = f(P^*B^*(P^{-1})^*) = P^*f(B^*)(P^{-1})^*,$$

and accordingly

$$BH = Hf(B^*), \tag{7}$$

where

$$H = PP^*. \tag{8}$$

The matrix H, defined by (8), is a positive definite hermitian matrix. It is now natural to make the following definition: *if H is any non-singular hermitian matrix, and, if $BH = Hf(B^*)$, the matrix B is normal with respect to H.*

---

[3] Wintner, *op. cit.*, p. 22.

A matrix normal with respect to the unit matrix is therefore a normal matrix. The matrix $f(B^*)$ is commutative with B^*. Further, if C is commutative with every matrix, which is commutative with B^*, C is a polynomial in B.[4] Hence we have the alternative definition; *if* $BH = HC$, *where* C *is commutative with every matrix, which is commutative with* B^*, *the matrix* B *is normal with respect to the non-singular hermitian matrix* H.

Let B_1 be any matrix similar to B so that

$$B = R^{-1}B_1R.$$

Then,

$$R^{-1}B_1RH = HR^*f(B^*_1)(R^*)^{-1}$$

and therefore

$$B_1H_1 = H_1f(B^*_1),$$

where

$$H_1 = RHR^*. \tag{9}$$

Consequently we have

LEMMA 2. *If* B *is normal with respect to* H *and* $B = R^{-1}B_1R$, *then* B_1 *is normal with respect to* $H_1 = RHR^*$. *Moreover, if* $BH = Hf(B^*)$, $B_1H_1 = H_1f(B^*_1)$.

Accordingly the theory of matrices, normal with respect to an hermitian matrix H, is similar to that of matrices, normal with respect to any matrix H_1, which satisfies (9) and which is therefore equivalent to H under a non-singular conjunctive transformation. In particular the theory of matrices normal with respect to a positive definite hermitian matrix is similar to that of normal matrices.

The canonical form of a normal matrix under unitary transformations is exceedingly simple; in fact, as remarked earlier, it is a diagonal matrix. Since a unitary matrix is a conjunctive automorph of the unit matrix, representative of all positive definite hermitian matrices, it is interesting to consider the corresponding problem for matrices normal with respect to an hermitian matrix H, which is not necessarily positive definite. Therefore, in what follows we discuss the problem: what are the possible canonical forms for a matrix B, normal with respect to an hermitian matrix H, under similarity transformations by matrices, which are conjunctive automorphs of H; i. e. what are the possible canonical forms for a matrix

$$C = RBR^{-1} \tag{10}$$

---

[4] Turnbull and Aitken, *Canonical Matrices*, Blackie and Sons (1932), p. 150.

where

(11) $$RHR^* = H$$

and

$$BH = Hf(B^*).$$

For brevity we shall call two matrices B and C, which satisfy (10) and (11), *H-equivalent.* It is therefore an immediate consequence of Lemma 2 that, if B is normal with respect to H and if B and C are H-equivalent, C is also normal with respect to H.

From the remarks following Lemma 2 it is apparent that it is not to be hoped to obtain satisfactory results, when H is a general hermitian matrix, but only when H is a suitably chosen representative of a class of conjunctively equivalent hermitian matrices. This is in fact what happens and we therefore only determine canonical forms of matrices normal with respect to H under H-equivalent transformations when H is suitably chosen. From the canonical forms we immediately deduce necessary and sufficient conditions that two matrices normal with respect to H be H-equivalent. The particular cases, when $f(A^*) = A^*$ or $f(A^*) = (A^*)^{-1}$, which correspond respectively, in the theory of normal matrices, to A being hermitian or unitary, are considered in section 5. These two cases have already been treated separately as quite unrelated problems. However, the methods employed in their solution are very similar and it was this similarity that suggested the discussion of the more general problem now under consideration.

1. Let A_1 and A_2 be two similar matrices, which are both normal with respect to the same non-singular hermitian matrix H. Then A_1 and A_2 are both similar to the same matrix Q and there therefore exist two non-singular matrices P_1 and P_2 such that

(12) $$A_i = P_i^{-1}QP_i, \qquad (i = 1, 2),$$

and

(13) $$S_i = P_iHP^*_i, \qquad (i = 1, 2).$$

Further by Lemma 2 the matrix Q is normal with respect to both of the hermitian matrices S_1 and S_2. We now prove a theorem which greatly simplifies our problem.

THEOREM 1. *Necessary and sufficient conditions that two matrices A_1 and A_2, both normal with respect to H, be H-equivalent is that there exist a non-singular matrix K such that*

(14) $$KQ = QK \text{ and } KS_1K^* = S_2,$$

where Q, S_1 *and* S_2 *are any matrices, which satisfy* (12) *and* (13).

First, let A_1 and A_2 be H-equivalent. Then there exists a non-singular matrix R satisfying

(15) $$RA_1R^{-1} = A_2$$

and

(16) $$RHR^* = H.$$

Let $K = P_2RP_1^{-1}$. Then

$$\begin{aligned} KQ &= P_2RP_1^{-1}Q = P_2RA_1P_1^{-1} \text{ by } (12),\\ &= P_2A_2RP_1^{-1} \text{ by } (15),\\ &= QP_2RP_1^{-1} \text{ by } (12),\\ &= QK. \end{aligned}$$

Further,

$$\begin{aligned} KS_1K^* &= P_2RP_1^{-1}S_1(P^*_1)^{-1}R^*P^*_2 = P_2RHR^*P^*_2 \text{ by } (13),\\ &= P_2HP^*_2 = S_2 \text{ by } (16) \text{ and } (13). \end{aligned}$$

Conversely, if a matrix K exists, which satisfies (14), and if $R = P_2^{-1}KP_1$, R satisfies (15) and (16) and consequently A_1 and A_2 are H-equivalent.

Hence in determining the possible canonical forms of a matrix A, normal with respect to H, under H-equivalent transformations we may proceed as follows. First we let Q be a suitably chosen canonical form of A under similarity transformations and let S be the corresponding hermitian matrix, conjunctively equivalent to H, with respect to which Q is normal. Then we determine a canonical form for S under conjunctive transformations by matrices which are commutative with Q. In other words we need only determine a canonical matrix W such that

(17) $$KSK^* = W \text{ and } KQ = QK.$$

We shall call such a transformation (17) by the matrix K an *admissible* transformation.

2. As mentioned at the end of the previous section we may take Q to be any matrix similar to A. We therefore choose Q to be the *diagonal block* matrix

(18) $$Q = [Q_1, Q_2, \cdots, Q_k],$$

where the latent roots of Q_i are all equal to λ_i, and $\lambda_i \neq \lambda_j$, if $i \neq j$. Since A is normal with respect to H, by Lemma 2, Q is normal with respect to S, so that

(19) $$QS = Sf(Q^*) = SM,$$

where M is the diagonal block matrix

$$M = [M_1, M_2, \cdots, M_k]$$

and

$$M_i = f(Q^*{}_i), \qquad (i = 1, 2, \cdots, k).$$

Let

$$S = (S_{ij}), \qquad (i, j = 1, 2, \cdots, k),$$

be a *partition* of S *similar* to that of Q in (18); i. e. S_{ij} is a matrix with the same number of rows as Q_i and the same number of columns as Q_j. It now follows from (19) that

$$(20) \qquad Q_i S_{ij} = S_{ij} M_j, \qquad (i, j = 1, 2, \cdots, k).$$

The latent roots of M_j, being $f(\bar{\lambda}_j)$, are all the same. Hence, either no latent root of M_j is the same as a latent root of Q_i or all latent roots of M_j are the same as those of Q_i. Since, when $i \neq j$, $\lambda_i \neq \lambda_j$, for a fixed value of i in (20) there is at most one value of j, for which the latent roots of M_j are the same as those of Q_i. But there must be at least one value of j, for which the latent roots of M_j coincide with those of Q_i, as otherwise S_{ij} would be zero for all values of j and S would be singular. There are therefore only two possibilities:

I. The latent roots of M_i are the same as the latent roots of Q_i, $S_{ij} = 0$ when $j \neq i$, and

$$(21) \qquad Q_i S_{ii} = S_{ii} M_i.$$

Since S is non-singular the matrix S_{ii} in (21) is non-singular.

II. The latent roots of M_j are the same as those of Q_i but i is different from j. Then $S_{ih} = 0$, if $h \neq j$. Since S is hermitian

$$S_{hi} = S^*{}_{ih} = 0, \qquad (h \neq j);$$

and since S is non-singular S_{ij} and S_{ji} are both non-singular. In place of (21) there are the two equations

$$Q_i S_{ij} = S_{ij} M_j \quad \text{and} \quad Q_j S_{ji} = S_{ji} M_i.$$

Since S_{ij} and S_{ji} are non-singular the latent roots of M_j are all equal to λ_i and those of M_i are all equal to λ_j.

Accordingly after a re-arrangement of the rows and the same re-arrangement of the columns of the matrices S, M and Q, we see that S becomes a diagonal block matrix. The blocks are of two distinct types;

$$\text{I.} \qquad S_{ii} = S^*{}_{ii}, \quad Q_i S_{ii} = S_{ii} M_i, \quad M_i = f(Q^*{}_i)$$

and

II. $$\begin{pmatrix} 0 & S_{ij} \\ S_{ji} & 0 \end{pmatrix}, \quad \begin{pmatrix} Q_i & 0 \\ 0 & Q_j \end{pmatrix}\begin{pmatrix} 0 & S_{ij} \\ S_{ji} & 0 \end{pmatrix} = \begin{pmatrix} 0 & S_{ij} \\ S_{ji} & 0 \end{pmatrix}\begin{pmatrix} M_i & 0 \\ 0 & M_j \end{pmatrix},$$
$$S_{ij} = S^*_{ji}.$$

It so happens that blocks of type II are very much simpler than those of type I. Accordingly we first consider the reduction of type II. Let

$$X = \begin{pmatrix} E & 0 \\ 0 & S_{ji}^{-1} \end{pmatrix},$$

where E is the unit matrix of the same order as S_{ij}. Then, since

$$X[Q_i, Q_j]X^{-1} = [Q_i, S_{ji}^{-1} Q_j S_{ji}] = [Q_i, M_i] = [Q_i, f(Q^*_i)],$$

$[Q_i, Q_j]$ is similar to $[Q_i, f(Q^*_i)]$. Further

$$X\begin{pmatrix} 0 & S_{ij} \\ S_{ji} & 0 \end{pmatrix}X^* = \begin{pmatrix} 0 & E \\ E & 0 \end{pmatrix}.$$

Since, in Theorem 1, Q is any matrix similar to A, in Q we may replace $[Q_i, Q_j]$ by $[Q_i, f(Q^*_i)]$. If this is done, S is a diagonal block matrix and the block $\begin{pmatrix} 0 & E \\ E & 0 \end{pmatrix}$ takes the place of the block $\begin{pmatrix} 0 & S_{ij} \\ S_{ji} & 0 \end{pmatrix}$. Hence we have

RESULT a. *If A is normal with respect to H, so that $AH = Hf(A^*)$, and, if there are two latent roots λ_i and λ_j of A, such that $\lambda_j = f(\bar{\lambda}_i)$, then $\lambda_i = f(\bar{\lambda}_j)$. The diagonal block matrix Q contains the block $[Q_i, f(Q^*_i)]$ and the corresponding block of S is* $\begin{pmatrix} 0 & E \\ E & 0 \end{pmatrix}$.

If λ_i is any latent root of A and $\lambda_i \neq f(\bar{\lambda}_i)$, there must exist a latent root λ_j of A such that

(22) $$\lambda_j = f(\bar{\lambda}_i), \quad \lambda_i = f(\bar{\lambda}_j).$$

Hence, if, for no latent root λ_h of A, is $\lambda_h = f(\bar{\lambda}_h)$, corresponding to each latent root λ_i of A there is a latent root λ_j, such that (22) is true. Then, as all blocks of Q are of type II, it follows that A is of order $2m$ and from result a that H has signature zero. By re-arranging the order of the blocks Q_i in (18) we obtain a matrix F, which is similar to Q and has the form

$$F = [F_1, f(F^*_1)],$$

where F_1 is a square matrix of order m. If Q is replaced by F, and this, by Theorem 1, is always possible,

$$S=\begin{pmatrix}0 & E_m\\ E_m & 0\end{pmatrix},$$

where E_m is the unit matrix of order m. We have therefore proved

THEOREM 2. *Let A be normal with respect to H and let $AH=Hf(A^*)$. If $\lambda_i \neq f(\bar{\lambda}_i)$ for all latent roots λ_i of A, then there exists a non-singular matrix R such that $RAR^{-1}=[F_1, f(F^*{}_1)]$ and $RHR^*=\begin{pmatrix}0 & E_m\\ E_m & 0\end{pmatrix}$. The matrices A and H are accordingly both of even order and the signature of H is zero.*

The matrix F_1 is not unique but may be replaced by any matrix similar to it. If, however, F_1 is taken in the classical canonical form, the matrix $[F_1, f(F^*{}_1)]$ is uniquely determined by A. Accordingly, if A is normal with respect to $S=\begin{pmatrix}0 & E_m\\ E_m & 0\end{pmatrix}$ and, if no latent root λ_i of A satisfies the equation $f(\bar{\lambda}_i)=\lambda_i$, A is S-equivalent to a unique canonical matrix $[F, f(F^*{}_1)]$. As a consequence we have

COROLLARY 1. *If A satisfies the hypotheses of Theorem 2 and A is similar to the matrix B, which is also normal with respect to H, then A and B are H-equivalent.*

3. Before considering a further reduction of the blocks of type I we prove three lemmas, the first of which is

LEMMA 3. *Let $Q=[Q_1, Q_2]$, $M=[M_1, M_2]$, $S=(S_{ij})$, $(i, j=1, 2)$, be similar partitions of three matrices Q, M and S, which satisfy* (19). *If S is hermitian and S_1 is non-singular, there exists a non-singular matrix K such that $KQ=QK$ and $KSK^*=[S_{11}, \sigma_{22}]$.*

As a consequence of (20)

$$S_{21}S_{11}^{-1}Q_1=S_{21}M_1S_{11}^{-1}=Q_2S_{21}S_{11}^{-1}.$$

Hence, if $K=\begin{pmatrix}E_1 & 0\\ -S_{21}S_{11}^{-1} & E_2\end{pmatrix}$, where E_i is the unit matrix of the same order as Q_i, $KQ=QK$. Further,

$$KSK^*=\begin{pmatrix}E_1 & 0\\ -S_{21}S_{11}^{-1} & E_2\end{pmatrix}\begin{pmatrix}S_{11} & S_{12}\\ S_{21} & S_{22}\end{pmatrix}\begin{pmatrix}E_1 & S_{11}^{-1}S_{12}\\ 0 & E_2\end{pmatrix}=\begin{pmatrix}S_{11} & 0\\ 0 & S_{22}-S_{21}S_{11}^{-1}S_{12}\end{pmatrix},$$

and the lemma is proved.

Let U be the auxiliary unit matrix of order n and V the auxiliary unit matrix of order m.[5] We now consider the matrix equation

$$\phi(U)D = D\psi(V), \tag{23}$$

where

$$\phi(x) = \sum_{\alpha=1}^{n-1} b_\alpha x^\alpha, \qquad \psi(x) = \sum_{\beta=1}^{m-1} c_\beta x^\beta \text{ and } b_1 c_1 \neq 0.$$

If $D = (d_{ij})$ $(i = 1, 2, \cdots, n;\ j = 1, 2, \cdots, m)$, (23) implies

$$\sum_{\alpha=1}^{n-1} b_\alpha d_{\alpha+i,j} = \sum_{\beta=1}^{m-1} c_\beta d_{i,j-\beta}, \qquad (i = 1, 2, \cdots, n;\ j = 1, 2, \cdots, m). \tag{24}$$

It is of course to be understood that in (24) $d_{\alpha+i,j} = 0$, if $\alpha + i > n$, and that $d_{i,j-\beta} = 0$, if $j - \beta \leq 0$.

Let us now suppose that $d_{rs} = 0$ for all values of r and s, for which $s - r \leqq k$. Then, if $j - i = k + 2$, equation (24) becomes

$$b_1 d_{1+i,j} = c_1 d_{i,j-1}, \tag{25}$$

or, when $j = 1$,

$$b_1 d_{1+i,1} = c_1 d_{i0} = 0,$$

so that $d_{1+i,1} = 0$. However (25) is valid only when $i > 0$ and, therefore, if $j = 1$, only when $j - i \leq 0$, i. e. when $k \leq -2$. Hence, if $k \leq -2$, $d_{rs} = 0$ for all values of r and s, for which $s - r \leq k + 1$. On the other hand, if $i = n$, (25) becomes

$$c_1 d_{n,j-1} = b_1 d_{n+1,j} = 0,$$

so that $d_{n,j-1} = 0$. Since $j - n = k + 2$ and $j \leq m$, $k \leq m - n + 2$. Hence, if $k \leq m - n - 2$, $d_{rs} = 0$ for all values of r and s, for which $s - r \leq k + 1$. Therefore, by induction, $d_{rs} = 0$, if $s - r \leq \{\text{maximum of } m - n - 1 \text{ and } -1\}$. When $m > n$, $0 \leq m - n - 1$, so that $d_{11} = 0$ and the first column of D is zero; when $m < n$, $m - n \leq -1$ and the last row of D is zero. We have therefore

LEMMA 4. *If D is a matrix satisfying* (23), *the first column of D is zero, when $m > n$, and the last row is zero, when $m < n$. If $m = n$, D is non-singular if, and only if d_{nn} is different from zero.*

If D is a square matrix, so that $m = n$, it follows from the above considerations that,

$$D = D_0 + D_1 + D_2 + \cdots + D_{n-1}, \tag{26}$$

---

[5] Turnbull and Aitken, *op. cit.*, p. 142.

where the only non-zero elements of D_j are those d_{rs} for which $s - r = j$. Equation (23) may now be written in the form

$$\sum_{\alpha=1}^{n-1} b_\alpha U^\alpha \sum_{j=0}^{n-1} D_j = \sum_{j=0}^{n-1} D_j \sum_{\beta=1}^{n-1} c_\beta U^\beta. \tag{27}$$

Equation (27) and the nature of the matrices D_j imply

$$\sum_{\alpha=1}^{s} b_\alpha U^\alpha D_{s-\alpha} = \sum_{\alpha=1}^{s} c_\alpha D_{s-\alpha} U^\alpha, \qquad (s = 1, 2, \cdots, n-1). \tag{28}$$

If $D_0 = \epsilon E$ and $D_1 = D_2 = \cdots = D_{r-1} = 0$, it is a consequence of the first r of the equations (28) that

$$b_i = c_i, \qquad (i = 1, 2, \cdots, r). \tag{29}$$

If in addition $b_{r+1+j} = c_{r+j} = 0$ for all $j > 0$, the $(r+1)$-th equation in (28) becomes

$$b_1 U D_r + \epsilon b_{r+1} U^{r+1} = b_1 D_r U. \tag{30}$$

When (30) is satisfied, the remaining of the equations (28) may be solved successively for $D_{r+1}, \cdots, D_{n-1}$.[6] We have therefore

LEMMA 5. *If equation* (30) *is satisfied there exists a matrix*

$$D = \epsilon E + D_r + D_{r+1} + \cdots + D_{n-1}$$

such that

$$D \sum_{\alpha=1}^{r-1} b_\alpha U^\alpha D^{-1} = \sum_{\alpha=1}^{r+1} b_\alpha U^\alpha.$$

4. The matrices in a block of type I satisfy (21) where the latent roots of Q_i all have the same value λ_i. Accordingly we temporarily drop all suffixes i or, what is equivalent to this, assume that all the latent roots of A have the same value λ. We may therefore take Q in the classical canonical form

$$Q = [Q_1, Q_2, \cdots, Q_k],$$

where

$$Q_i = \lambda(E_i + U_i), \lambda \neq 0; \qquad Q_i = U_i, \lambda = 0. \tag{31}$$

In (31) E_i is the unit matrix of order e_i and U_i the auxiliary unit matrix of the same order. The elementary divisors of $A - xE$ are therefore $(x-\lambda)^{e_i}$, $(i = 1, 2, \cdots, k)$. We further suppose that $e_1 \geqq e_2 \geqq \cdots \geqq e_k$. The matrix $M = [M_1, M_2, \cdots, M_k]$, where

---

[6] H. W. Turnbull, "Power vectors," *Proceedings of the London Mathematical Society*, Series 2, vol. 39 (1934), part 2, pp. 106-146.

$$(32) \qquad M_i = \lambda(E_i + \sum_{j=1}^{e_i-1} a_j U'^{\,j}_i),\ \lambda \neq 0; \qquad M_i = \sum_{j=1}^{e_i-1} a_j U'^{\,j}_i,\ \lambda = 0.$$

Since $M_i = f(Q^*_i)$, the a_j in (32) are determined uniquely by the polynomial $f(x)$ and are independent of i.

Let

$$S = (S_{ij}), \qquad (i, j = 1, 2, \cdots, k),$$

be a partition of S similar to that of Q. Then as a consequence of (20) we have

$$(33) \qquad U_i S_{ij} = S_{ij} \sum_{k=1}^{e_j-1} a_k U'^{\,k}_j, \qquad (i, j = 1, 2, \cdots, k).$$

Equations (33) are the same whether λ is or is not zero. If T_j is the secondary unit matrix of order e_j [7]

$$T_j^2 = E_j \quad \text{and} \quad T_j U'_j = U_j T_j.$$

Hence, if

$$T = [T_1, T_2, \cdots, T_k]$$

and

$$S = DT = (D_{ij} T_j), \qquad (i, j = 1, 2, \cdots, k),$$

equations (33) become

$$U_i D_{ij} T_j = D_{ij} T_j \sum_{k=1}^{e_j-1} a_k U'^{\,k}_j = D_{ij} \sum_{k=1}^{e_j-1} a_k U_j^k T_j$$

or

$$(34) \qquad U_i D_{ij} = D_{ij} \sum_{k=1}^{e_j-1} a_k U_j^k.$$

Since M is similar to Q, $a_1 \neq 0$ and equations (34) are all of the type (23). If $e_1 > e_2$, by Lemma 4 the last row of D_{1j} is zero except when $j = 1$. Since S is non-singular D is non-singular and the last row of D_{11} is not zero. The only element in the last row of D_{11}, which is different from zero, is the element in the last column and therefore, by Lemma 4, D_{11} is non-singular. If $e_1 = e_2 = \cdots = e_c > e_{c+1}$ and D_{ii} is non-singular for some value of i, $1 \leq i \leq c$, by a suitable re-arrangement of the rows of D and the same re-arrangement of the columns we may move D_{ii} into the place of D_{11} without disturbing Q or M. Accordingly we may suppose in this case that D_{11} is non-singular. There only remains the case, in which D_{ii} is singular for all values of i, $1 \leq i \leq c$. Let d_{ij} be the element in the last row and the last column of D_{ij}. Then

$$(35) \qquad d_{ii} = 0, \qquad (i = 1, 2, \cdots, c).$$

---

[7] Turnbull and Aitken, *op. cit.*, p. 11.

Since D is non-singular, for at least one value of i, $1 < i \leq c$, $d_{1i} \neq 0$, as otherwise one row of D would be zero. Hence without any loss of generality we may assume that $d_{12} \neq 0$. Let E be the unit matrix of order e_1 and ϵ be a complex number. Then

$$\begin{pmatrix} E & \epsilon E \\ 0 & E \end{pmatrix}\begin{pmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{pmatrix}\begin{pmatrix} E & \epsilon E \\ 0 & E \end{pmatrix}^{*} = \begin{pmatrix} K_{11} & K_{12} \\ K_{21} & K_{22} \end{pmatrix}\begin{pmatrix} T_1 & 0 \\ 0 & T_1 \end{pmatrix},$$

where $K_{11} = D_{11} + \epsilon D_{21} + \bar{\epsilon} D_{12} + \epsilon\bar{\epsilon} D_{22}$. The element in the last row and last column of K_{11} is $k_{11} = d_{11} + \epsilon d_{21} + \bar{\epsilon} d_{12} + \epsilon\bar{\epsilon} d_{22} = \epsilon d_{21} + \bar{\epsilon} d_{12}$ by (35). Since ϵ is arbitrary, for at least one value of ϵ, $k_{11} \neq 0$. The matrix $\begin{pmatrix} E & \epsilon E \\ 0 & E \end{pmatrix}$ is commutative with $[Q_1, Q_2]$ and therefore the above transformation is admissible. Hence by Lemma 4, if $k_{11} \neq 0$, K_{11} is non-singular. We may suppose then that such a transformation, if necessary, has been made, and that D_{11} and therefore S_{11} is non-singular. By Lemma 3 there is therefore an admissible transformation which reduces S to the form $[S_{11}, \sigma]$. By repetitions of the above process we finally reduce S by an admissible transformation to the diagonal block form

$$S = [S_1, S_2, \cdots, S_k],$$

where S_i is of the same order as Q_i and

$$Q_i S_i = S_i f(Q^*_i) = S_i M_i. \tag{36}$$

If $A - xE$ has the single elementary divisor $(x - \lambda)^n$, (19) coincides with (33). Hence we need now only consider this particular case, in which

$$Q = \lambda(E + U),\ \lambda \neq 0; \qquad Q = U,\ \lambda = 0,$$
$$M = \lambda(E + \sum_{j=1}^{n-1} a_j U'^{j}),\ \lambda \neq 0; \qquad M = \sum_{j=1}^{n-1} a_j U'^{j},\ \lambda = 0,$$

and

$$S = DT$$

where D is given by (26). The equations that correspond to (24) are

$$d_{i+1,j} = \sum_{\beta=1}^{n-1} a_\beta d_{i,j-\beta}$$

and in particular, when $j = i + 1$,

$$d_{i+1,i+1} = a_1 d_{ii}. \tag{37}$$

Since S is hermitian, $D_j T$ is hermitian and in particular so is $D_0 T$. Therefore

$$d_{n-i,n-i} = \bar{d}_{i+1,i+1}. \tag{38}$$

As a consequence of (37) and (38) a_1 is of unit modulus. If $n = 2m + 1$ and $i = m$, (38) becomes

$$d_{m+1,m+1} = \bar{d}_{m+1,m+1} = \epsilon\rho^2$$

where ρ is real and $\epsilon = \pm 1$. It now follows from (37) that

$$D_0 = \epsilon\rho^2[a_1{}^{-m}, a_1{}^{-(m-1)}, \cdots, a_1{}^{-1}, 1, a_1, \cdots, a_1{}^{m-1}, a_1{}^{m}].$$

Let b be a particular one of the square roots of a_1 and K be the matrix

$$K = \rho^{-1}[b^m, b^{m-1}, \cdots, b, 1, b^{-1}, \cdots, b^{-m}].$$

Then, since $b = \sqrt{a_1}$, and $|a_1| = 1$, $b^{-1} = \bar{b}$ and consequently $KT = TK^*$. Therefore,

$$KD_0TK^* = KD_0KT = \epsilon T, \qquad \epsilon = \pm 1, \tag{39}$$

while a simple calculation shows that

$$KUK^{-1} = bU. \tag{40}$$

If, however, $n = 2m$, (37) and (38) imply that $\bar{d}_{mm} = d_{m+1,m+1}$. Hence, if $d_{m+1,m+1} = \rho^2 e^{i\theta}$, where ρ is real, $a_1 = e^{-2i\theta}$. Since $e^{i\theta}$ is determined by d_{mm}, if b is a definite square root of a_1, $e^{i\theta} = \epsilon b$, $\epsilon = \pm 1$. Therefore

$$D_0 = \epsilon\rho^2[b^{-(2m-1)}, b^{-(2m-3)}, \cdots, b^{-1}, b, \cdots b^{(2m-3)}, b^{(2m-1)}].$$

If

$$K = \rho^{-1}[b^{(2m-1)/2}, \cdots, b^{\frac{1}{2}}, b^{-\frac{1}{2}}, \cdots, b^{-[(2m-1)/2]}],$$

equations (39) and (40) are again valid. Moreover as a consequence of (40),

$$(K^{-1})^*U'K^* = \bar{b}U' = U'/b. \tag{41}$$

We now prove

LEMMA 6. *There exists a non-singular matrix R such that*

$$RUR^{-1} = \sum_{i=1}^{n-1} b_iU^i \quad \text{and} \quad RSTR^* = \epsilon T,$$

where $b_1 = b$ and $\epsilon = \pm 1$.

We shall prove this lemma by induction and therefore assume that there exists a matrix W such that

$$WUW^{-1} = \sum_{i=1}^{r} b_iU^i \quad \text{and} \quad WSW = DT, \tag{42}$$

where D is given by (26) and

$$D_0 = \epsilon D, \qquad D_1 = D_2 = \cdots = D_{r-1} = 0. \tag{43}$$

We first note that the matrix K, which satisfies (39) and (40), is a matrix W satisfying (42), when $r = 1$. Since $US = SM$, where $M = \sum_{j=1}^{n-1} a_j U'^j$,

$$WUW^{-1}WSW^* = WSW^*(W^*)^{-1}MW$$

and by (42)

$$\sum_{i=1}^{r} b_i U^i DT = DT \sum_{i=1}^{n-1} c_i U'^i$$

or finally

$$\sum_{i=1}^{r} b_i U^i D = D \sum_{i=1}^{n-1} c_i U^i. \tag{44}$$

Equation (44) is of the form (23) and it follows from (43) that (29) is true and

$$b_1 U D_r = b_1 D_r U + \epsilon c_{r+1} U^{r+1}. \tag{45}$$

If $b_{r+1} = c_{r+1}/2$, (44) becomes

$$b_1 U \frac{-D_r}{2} + \epsilon b_{r+1} U^{r+1} = b_1 \frac{-D_r}{2} U.$$

This last equation is the same as (30), if D_r is replaced by $-D_r/2$. Therefore by Lemma 5 we can determine matrices $F_{r+1}, F_{r+2}, \cdots, F_{n-1}$, where F_j has the same form as D_j, such that the matrix

$$N = \epsilon E - D_r/2 + F_{r+1} + \cdots + F_{n-1}$$

satisfies

$$N \sum_{i=1}^{r} b_i U^i N^{-1} = \sum_{i=1}^{r+1} b_i U^i. \tag{46}$$

Since DT is hermitian $D_r T = TD^*_r$, and therefore

$$\begin{aligned} NDTN^* = {} & (\epsilon E - D_r/2 + F_{r+1} + \cdots + F_{n-1}) \\ & \times D(\epsilon E - D_r/2 + G_{r+1} + \cdots + G_{n-1})T, \end{aligned} \tag{47}$$

where G_j has the same form as D_j. On multiplying out the right hand side of (47) we find that

$$NDTN^* = D_0 + H_{r+1} + \cdots + H_{n-1}, \tag{48}$$

where H_j is of the same form as D_j. Equations (42), (46) and (48) imply that (42) and (43) are true when r is replaced by $r+1$ and W by NW. Our induction is now complete and the lemma proved.

By (45)

$$(UD_r - D_rU)T = \frac{\epsilon c_{r+1}}{b_1} U^{r+1}T.$$

Since D_rT is hermitian,

$$\{(UD_r - D_rU)T\}^* = TD^*_rU' - TU'D^*_r = D_rUT - UD_rT$$
$$= -(UD_r - D_rU)T.$$

Hence $\frac{\epsilon c_{r+1}}{b} T$ is anti-hermitian and therefore c_{r+1}/b is a pure imaginary quantity. Since $b_{r+1} = c_{r+1}/2$ we have

(49) $$b_{r+1} = ib_1\beta_{r+1},$$

where β_{r+1} is real. Further, since the c_i in (44) are determined uniquely, by the polynomial $f(x)$, in terms of $b_1, b_2, \cdots, b_r$ it follows that the b_j in Lemma 6 are determined uniquely, apart from the sign of b_1, by the polynomial $f(x)$. If R is the matrix of Lemma 6,

$$RQR^{-1} = \lambda(E + \sum_{i=1}^{n-1} b_iU^i),\ \lambda \neq 0, \qquad RQR^{-1} = \sum_{i=1}^{n-1} b_iU^i,\ \lambda = 0,$$
$$(R^*)^{-1}MR^* = \lambda(E + \sum_{i=1}^{n-1} b_iU^{i'}),\ \lambda \neq 0, \qquad (R^*)^{-1}MR^* = \sum_{i=1}^{n-1} b_iU^{i'},\ \lambda = 0,$$
$$RSR^* = \epsilon T.$$

We have now proved

Result b. *Let A be normal with respect to H, so that $AH = Hf(A^*)$ and let λ_j be a latent root of A such that $f(\bar{\lambda}_j) = \lambda_j$. If $(x - \lambda_j)^{e_k}$ is an elementary divisor of $A - xE$, the matrix A is similar to a diagonal block matrix Q, which contains the block*

(50) $$Q_{jk} = \lambda_j(E_k + b_{1j}U_k + ib_{1j} \sum_{r=2}^{e_k-1} \beta_{rj}U_k^r),\ \lambda_j \neq 0,$$
$$= b_{1j}U_k + ib_{1j} \sum_{j=2}^{e_k-1} \beta_{rj}U_k^r,\ \lambda_j = 0.$$

The corresponding block in the matrix S, with respect to which Q is normal, is $\epsilon_{jk}T_k$, where $\epsilon_{jk} = \pm 1$. In (50) b_{1j} is of modulus one and each β_{rj} is real.

By combining results a and b we have

Theorem 3. *If A is normal with respect to H, there exists a non-singular matrix P, such that $PAP^{-1} = Q$ and $PHP^* = S$, where Q and S are diagonal* *block matrices. Corresponding to each pair of latent roots of type II Q contains a block $[Q_i, f(Q^*_i)]$ given in result a and S the block $\begin{pmatrix} 0 & E \\ E & 0 \end{pmatrix}$; corresponding to each elementary divisor $(x-\lambda_j)^{e_k}$ of type I Q contains the block Q_{jk} given by (50) and S the block $\epsilon_{jk}T_k$.*

The matrices Q and S of Theorem 3 are canonical forms for A and H. Apart from the ϵ_{jk}, which have the value ± 1, everything in the canonical forms is uniquely determined by A, H and $f(x)$. If $(x-\lambda_j)^{e_k}$ occurs exactly t times among the elementary divisors of $A-xE$, with this elementary divisor are associated $t\ \epsilon_{jk}$ each of which has the value ± 1. The number of positive ϵ_{jk} is called the *index* associated with the elementary divisor $(x-\lambda_j)^{e_k}$. We now justify this terminology by showing that the index associated with each elementary divisor, as defined above, is uniquely determined by the matrices A and H.

Let Q, S_1 and Q, S_2 be two canonical forms for A and H. Then by Theorem 1, there exists a non-singular matrix K commutative with Q such that $KS_1K^* = S_2$. Since K is commutative with Q, K is a diagonal block matrix partitioned similarly to Q in (18). Consequently we need only consider the case in which all latent roots of Q are the same and of type II, since S_1 coincides with S_2 as far as blocks of type I are concerned. Let

$$Q = [Q_1, Q_2, \cdots, Q_k],\ S_1 = [\epsilon_1 T_1, \epsilon_2 T_2, \cdots, \epsilon_k T_k],\ S_2 = [\rho_1 T_1, \rho_2 T_2, \cdots, \rho_k T_k],$$

where $\epsilon_i = \pm 1$, $\rho_i = \pm 1$ and Q_j is of the form (50). If

$$K = (K_{ij}), \qquad (i, j = 1, 2, \cdots, k),$$

is a partition of K similar to that of Q,

$$\sum_{a=1}^{k} K_{ia}\epsilon_a T_a K^*_{ja} = \delta_{ij}\rho_i T_i, \tag{51}$$

where δ_{ij} is the Kronecker δ. If k_{ij} is the element in the top left-hand corner of K_{ij} and $e_{c-1} > e_c = e_{c+1} = \cdots = e_d > e_{d+1}$, (51) implies[8]

$$\sum_{a=c}^{d} k_{ia}\epsilon_a \bar{k}_{ja} = \delta_{ij}\rho_i, \qquad (i, j = c, c+1, \cdots, d). \tag{52}$$

Since K is non-singular, $|k_{ij}| \neq 0$, $(i, j = c, c+1, \cdots, d)$ and by (52) $[\epsilon_c, \epsilon_{c+1}, \cdots, \epsilon_d]$ is conjunctively equivalent to $[\rho_c, \rho_{c+1}, \cdots, \rho_d]$. Hence the index associated with $(x-\lambda)^{e_c}$ in the canonical form Q, S_1 is the same as that in the form Q, S_2. Consequently we have

---

[8] John Williamson, "The equivalence of non-singular pencils of hermitian matrices in an arbitrary field," *American Journal of Mathematics*, vol. 57 (1935), pp. 484-485.

THEOREM 4. *If A and B are normal with respect to H, A and B are H-equivalent if, and only if, the two matrices $A-xE$ and $B-xE$ have the same elementary divisors and if the indices associated with each elementary divisor of type I are the same for both matrices.*

If H is positive definite each elementary divisor is of type I and is linear while each index has the value one. We have therefore the well known

COROLLARY 1. *Two normal matrices which are similar are equivalent under a unitary transformation.*

5. Special cases. Let $f(x)=x$, so that $f(A^*)=A^*$. A latent root of type I must now be real. If λ_j is real, we have from (50)

$$Q^*_{jk}=\lambda_j(E_k+\bar{b}_{1j}U'_k-i\bar{b}_{1j}\sum_{r=2}^{e_k-1}\beta_{rj}U'^{r}_k)=\lambda_j(E_k+b_{1j}U'_k+ib_{1j}\sum_{r=2}^{e_k-1}\beta_{rj}U'^{r}_k).$$

Therefore b_{1j} is real and $\beta_{rj}=0$. Since b_{1j} has modulus one $b_{1j}=\pm 1$ and can therefore be taken as $+1$. The matrix

$$Q_{jk}=\lambda_jE_k+\lambda_jU_k.$$

Since $AH=HA^*=(AH)^*$, the matrix $H_1=AH$ is hermitian. If $PAP^{-1}=Q$ and $PHP^*=S$, $PH_1P^*=S_1=QS$. The matrices S and $S_1=QS$ are therefore canonical forms for the pair of hermitian matrices H and H_1 under conjunctive transformations. From Theorem 4 we can therefore deduce necessary and sufficient conditions for the conjunctive equivalence of two pairs of hermitian matrices.[9] If $f(x)=-x$, a latent root λ_j of type I must be pure imaginary and $b_{1j}=i$, $\beta_{rj}=0$. This could of course have been deduced from the previous case, since if $A=A^*$, $(iA)^*=-iA$.

Let $f(A^*)=(A^*)^{-1}$, so that A is a conjunctive automorph of H. If a latent root λ_j is of type I, $\lambda_j=-1/\bar{\lambda}_j$, so that $\lambda_j\bar{\lambda}_j=1$ or $\lambda_j=e^{i\theta}$. On dropping all suffixes j we have from (50)

---

[9] H. W. Turnbull, "On the equivalence of pencils of hermitian forms," *Proceedings of the London Mathematical Society*, vol. 39 (1935), pp. 232-248; M. H. Ingraham and K. W. Wegner, "The equivalence of pairs of hermitian matrices," *Transactions of the American Mathematical Society*, vol. 38 (1935), pp. 145-162; G. R. Trott, "On the canonical form of a non-singular pencil of hermitian matrices," *American Journal of Mathematics*, vol. 56 (1936), no. 3, pp. 359-391. The methods used in Trott's paper are very similar to those of this paper.

$$Q_{jk} = e^{i\theta}(E + b_1 U + b_1 i \sum_{r=2}^{e-1} \beta_r U^r) \tag{53}$$

and

$$f(Q^*_{jk}) = (Q^*_{jk})^{-1} = e^{i\theta}(E + b_1 U' + b_1 i \sum_{r=2}^{e-1} \beta_r U'^r).$$

Therefore

$$(Q_{jk})^{-1} = e^{-i\theta}(E + \bar{b}_1 U - \bar{b}_1 i \sum_{r=2}^{e-1} \beta_r U^r). \tag{54}$$

On multiplying (53) and (54), the coefficient of U on the right is $b_1 + \bar{b}_1$ and on the left is zero. Hence $b_1 = i$ and we have

$$E = U^2 + (E - \sum_{r=2}^{e-1} \beta_r U^r)^2. \tag{55}$$

If $z = \sum_{r=2}^{e-1} \beta_r U^r$ and $U = x$, (55) may be written in the form

$$(1 - z)^2 = 1 - x^2. \tag{56}$$

Therefore

$$(1 - z) = (1 - x^2)^{\frac{1}{2}},$$

$$z = \tfrac{1}{2}U^2 + \tfrac{1}{8}U^4 + \cdots 1 \cdot 3 \cdot 5 \cdots (2k - 3) U^{2k}/2^k k! + \cdots. \tag{57}$$

Consequently (53) becomes

$$Q_{jk} = e^{i\theta}(E + iU - \tfrac{1}{2}U^2 - \tfrac{1}{8}U^4 - \cdots - 1 \cdot 3 \cdot 5 \cdots (2k - 3) U^{2k}/2^k k! + \cdots. \tag{58}$$

The canonical form obtained from Theorem 3 by giving Q_{jk} the value (58) serves as an alternative to one found in a previous paper for the particular case now under consideration.[10]

6. If A is normal with respect to H so that $AH = Hf(A^*)$, the polynomial $f(x)$ is, as remarked in the introduction, by no means a general polynomial. If A has no latent root of type I, the nature of $f(x)$ can be deduced from Theorem 2. We now consider the nature of $f(x)$ when $A - xE$ has the single elementary divisor $(\lambda - x)^n$. By (50)

$$Q = \lambda(E + b_1 U + b_1 i \sum_{r=2}^{n-1} \beta_r U^r)$$

and

$$f(Q^*) = \lambda(E + b_1 U' + b_1 i \sum_{r=2}^{n-1} \beta_r U'^r)$$

and

---

[10] John Williamson, "Quasi-unitary matrices," *Duke Mathematical Journal*, vol. 3 (1937), no. 4, pp. 720-722.

$$Q^* = \bar{\lambda}(E + \bar{b}_1 U' - \bar{b}_1 i \sum_{r=2}^{n-1} \beta_r U'^r).$$

If $\dfrac{Q^* - \bar{\lambda} E}{\bar{\lambda}\bar{b}_1} = \eta$,

(59)
$$\frac{f(Q^*) - \lambda E}{\lambda b_1} = 2U' - \eta$$

while

(60)
$$\eta = U' - i \sum_{r=2}^{n-1} \beta_r U'^r.$$

Equation (60) may be solved for U' as a power series in η by an inductive process and therefore by (59), $f(Q^*)$ is determined as a polynomial in Q^*. If all the latent roots of A have the same value λ, $f(Q^*)$ remains unchanged even if $A - xE$ has more than one elementary divisor. The general case follows from the above by replacing $\bar{\lambda}\bar{b}_1\eta$ by the principal nilpotent element of Q^* associated with λ.[11]

7. The matrix S in the canonical form of Theorem 3 is determined by A and H and not by H alone. The blocks of S are of two types

$$\text{Type I} \quad \epsilon T_j; \qquad\qquad \text{Type II} \quad \begin{pmatrix} 0 & E \\ E & 0 \end{pmatrix}.$$

It is possible to determine conjunctive transformations which reduce both types to diagonal form and therefore to find a matrix W such that

$$WSW^* = [E_s, -E_t],$$

where E_j is the unit matrix of order j and $s - t$ is the signature of H or S. The matrix

$$WQW^{-1} = R$$

can then be calculated. The matrices R thus obtained give unique canonical forms for all matrices normal with respect to $[E_s, -E_t]$ under similarity transformations by matrices which are conjunctive automorphs of $[E_s, -E_t]$. If t or s is zero, we obtain the result from which we started; every normal matrix can be reduced to diagonal form by a unitary transformation. We have not carried out this reduction as the canonical form R is not nearly as simple as that of Q in Theorem 3.[12]

THE JOHNS HOPKINS UNIVERSITY.

---

[11] Wedderburn, *op. cit.*, p. 29.

[12] Reductions of the type mentioned have been carried out in special cases. See (10) and John Williamson, "On the normal forms of linear canonical transformations in dynamics," *American Journal of Mathematics*, vol. 59 (1937), no. 3, pp. 614-617.

SUBRINGS OF DIRECT SUMS.*

By NEAL H. McCOY.

1. Introduction.

1. Introduction. If K is any ring, a direct sum of rings K is understood to be the ring of all functions with values in K, defined on some finite or infinite set M. In a recent paper[1] the following criterion was established for determining when a ring is isomorphic to a subring of a direct sum. *A necessary and sufficient condition that a ring R be isomorphic to a subring of a direct sum of rings K is that for every element $a \neq 0$ in R there exists a homomorphism h of R into a subring of K such that $h(a) \neq 0$.*

In the paper referred to, this theorem was used to show that if every element a of a commutative ring R satisfies the conditions $a^p = a$ and $pa = 0$, where p is a fixed prime, then R is isomorphic to a subring of a direct sum of Galois fields $GF(p)$. This is, in itself, a generalization of a theorem concerning Boolean rings, proved by Stone in a different manner.[2] The primary purpose of the present paper is to establish several extensions of these results as well as some other theorems of a related nature. The method used in establishing the existence of the necessary homomorphisms is believed to be somewhat simpler than those used previously.

2. Preliminary definitions. Let R be a given ring, and a any non-zero element of R. If there exists a positive integer n such that $na = 0$, the least such n will be called the *order* of a, this coinciding with the usual definition of order in the additive group of the ring. If every element of R is of order n, then n must clearly be a prime p, and we may say that R has the *characteristic*

---

* Received June 25, 1937.

[1] N. H. McCoy and Deane Montgomery, "A representation of generalized Boolean rings," to appear in the *Duke Mathematical Journal*. I am indebted to Professor Montgomery for several helpful suggestions during the preparation of the present paper.

Note added February 23, 1938. The paper just referred to has appeared in the *Duke Mathematical Journal*, vol. 3 (1937), pp. 455-459. In a more recent paper, which has been submitted to the same journal, it is shown that every commutative ring without nilpotent elements is isomorphic to a subring of a direct sum of fields. However, this general result does not give much or any information as to just what fields are involved in such a direct sum representation of a given ring. It will be noted that Theorems 2, 3 and 5 of the present paper furnish this information for the case of suitably restricted rings.

[2] M. H. Stone, "The theory of representations for Boolean algebras," *Transactions of the American Mathematical Society*, vol. 40 (1936), pp. 37-111.

p. If no element of R has finite order, we say that R has *characteristic* 0. In the sequel we shall refer to a ring of characteristic k, it being understood that k is either 0 or a prime p.

The commutative ring R will be called an *algebraic ring* if every element $a \neq 0$ of R satisfies an equation of the type

$$f_a(x) \equiv m_0x^n + m_1x^{n-1} + \cdots = 0 \qquad (m_0a^n \neq 0), \tag{1}$$

the coefficients m_i being integers. If R has no unit element, it is of course assumed that this equation has no constant term.

3. Imbedding theorem.

We now prove the following theorem.

THEOREM 1. *An algebraic ring R of characteristic k, without a unit element, may be imbedded in an algebraic ring R' of characteristic k with a unit element. If R contains no nilpotent elements, then R' contains no nilpotent elements.*

The proof is perhaps best divided into two cases according as $k = 0$ or $k = p$. Suppose first that $k = 0$. Consider the ring consisting of the pairs (a, n), where a is in R and n is an integer, with addition and multiplication defined as follows:[3]

$$(a, n) + (b, m) = (a + b, n + m),$$
$$(a, n)(b, m) = (ab + nb + ma, nm).$$

By $(a, n) = (b, m)$ we mean that $a = b$ and $n = m$. It follows at once that this is a ring of characteristic 0, with unit element $(0, 1)$, and with a subring isomorphic to R, namely the subring consisting of those elements of the form $(a, 0)$. It will be noted that this procedure shows the possibility of adjoining to R a ring element e having the following properties: (i) $ae = ea = a$ for all elements a of R, (ii) $e^2 = e$, and (iii) $a + ne = b + me$ implies $a = b$, $n = m$. The desired ring R' is then the ring $R[e]$ whose elements are of the form $a + ne$. Clearly e is the unit element of R'.

If $f(x) = 0$ is the equation (1) satisfied by a, then by Taylor's theorem we find that

$$f(a) = f(-ne) + f'(-ne)(a + ne) + (1/2!)f''(-ne)(a + ne)^2 + \cdots = 0.$$

Thus an arbitrary element $a + ne$ of R' satisfies an algebraic equation with integral coefficients, and R' is an algebraic ring. If R has no nilpotent elements, clearly no element $a + ne$ of R' can be nilpotent, and the theorem is established for this case.

---

[3] J. L. Dorroh, "Concerning adjunctions to algebras," *Bulletin of the American Mathematical Society*, vol. 38 (1932), pp. 85-88.

Suppose now that R is of characteristic p, and let the elements of $GF(p)$ be denoted by $\bar{0}, \bar{1}, \cdots, \overline{p-1}$. Consider the ring of pairs $(a, \bar{n})$, where a is in R and $\bar{n}$ is in $GF(p)$, with addition and multiplication defined as follows:

$$(a, \bar{n}) + (b, \bar{m}) = (a + b, \bar{n} + \bar{m}),$$
$$(a, \bar{n})(b, \bar{m}) = (ab + nb + ma, \bar{n}\bar{m}).$$

This ring is of characteristic p, with unit element $(0, \bar{1})$, and containing the subring of elements $(a, \bar{0})$ which is isomorphic to R. We may now proceed in a manner entirely analogous to that used above. The details will therefore be omitted.

4. Principal theorems on subrings of direct sums. Let P_k denote the prime field of characteristic k. Thus P_0 will be the field of rational numbers, and P_p will denote the $GF(p)$. Let C_k be the essentially unique algebraically closed, algebraic field in which P_k can be imbedded.[4] We shall now prove the following theorem.

THEOREM 2. *Every algebraic ring R of characteristic k without nilpotent elements is isomorphic to a subring of a direct sum of fields C_k.*

In view of the preceding theorem, it is sufficient to limit ourselves to the case in which R has a unit element e. Let I denote the subring of R consisting of the integral multiples of e. Thus if $k = 0$, I is isomorphic to the ring of rational integers, while if $k = p$, I is isomorphic to $P_p = GF(p)$. We may clearly consider the coefficients in $f_a(x)$ given by (1) as elements of I.

If D is any ring, we denote by $D[x]$ the ring of polynomials in an indeterminate x with coefficients in D. We now establish the following lemma.

LEMMA 1. *Let D be a subring of C_k, and M an ideal in $D[x]$ containing no non-zero element of D. Then there exists an element ρ of C_k such that $g(\rho) = 0$ provided $g(x) \equiv 0(M)$.*

Since D contains no divisors of zero, it may be imbedded in a quotient field $D' \subset C_k$.[5] Let $h(x)$ be an element of M of minimum degree ($\geqq 1$). If $g(x)$ is in M, then in $D'[x]$ we may write

$$g(x) = q(x)h(x) + r(x),$$

---

[4] See B. L. van der Waerden, *Moderne Algebra* I, p. 199.

[5] van der Waerden, *op. cit.*, p. 46.

the degree of $r(x)$ being less than the degree of $h(x)$. This equation may now be multiplied by a properly chosen non-zero element d of D such that $dq(x)$ and $dr(x)$ have coefficients in D. It then follows that $dr(x) \equiv 0(M)$, and being of degree less than the degree of $h(x)$ therefore vanishes identically. Thus

$$dg(x) = dq(x)h(x).$$

and for the ρ of the lemma we only need to choose an arbitrary root in C_k of the equation $h(x) = 0$.

Before proceeding to the proof of the theorem, we establish another lemma. If S is a subring of R, and a an element of R not in S, then by $S[a]$ we mean the ring generated by elements of S together with a, that is the ring of polynomials in a with coefficients in S.

LEMMA 2. *Let S be a subring of R containing the unit element e of R, and h a given homomorphism $S \rightarrow D$, where D is a subring of C_k. If a is an element of R not in S, then the homomorphism h may be extended*[6] *to a homomorphism $h' : S[a] \rightarrow D_1$, where $D \subset D_1 \subset C_k$.*

The elements of $S[a]$ are expressible as polynomials in a with coefficients in S. Suppose by the given homomorphism h, an arbitrary element s of S corresponds to the element $\bar{s}$ of D. To the element $\sum s_i a^i$ of $S[a]$ we may now make correspond the element $\sum \bar{s}_i x^i$ of $D[x]$. The set of all polynomials of $D[x]$ which thus correspond to the zero element of $S[a]$ forms an ideal M in $D[x]$. A polynomial $\sum \bar{s}_i x^i$ thus belongs to M if and only if there exist elements s_i of S such that $\sum s_i a^i = 0$ and $s_i \rightarrow \bar{s}_i$ by h. We proceed to show that M contains no non-zero element of D.

We wish to show that if $\sum_{k=0}^{m} s_k a^k = 0$, $\bar{s}_k = 0$ $(k \neq 0)$, then $\bar{s}_0 = 0$. For convenience, let us denote $\sum_{k=0}^{m} s_k a^k$ by A. From the equation (1) satisfied by a, it follows that

$$m_0 a^n = -(m_n e + \cdots + m_1 a^{n-1}).$$

Thus $m_0{}^2 a^{n+1}$, $m_0{}^3 a^{n+2}, \cdots$ are also expressible as linear combinations of $e, a, \cdots, a^{n-1}$, with integral coefficients. Hence if l is a positive integer $> m - n$, we may write

$$m_0{}^{l+i} a^i A = \sum_{k=0}^{m} s_k m_0{}^{l+i} a^{k+i} \qquad (i = 0, 1, \cdots, n-1),$$

---

[6] By this we mean that under h' the elements of S have the same image as under the given homomorphism h.

then replace $m_0{}^{l+i}a^{k+i}$ (for $k+i \geqq n$) by linear combinations of $e, a, \cdots, a^{n-1}$, and collect coefficients of the different powers of a. We are thus led to the following equations:

$$(2) \qquad m_0{}^{l+i}a^iA = \sum_{j=0}^{n-1} b_{ij}a^j = 0 \qquad (i = 0, 1, \cdots; n-1),$$

where the b_{ij} are linear combinations of the s_k with integral coefficients. From the method of formation of the b_{ij}, it follows that $b_{ii} = m_0{}^{l+i}s_0 + b'_{ii}$, where b'_{ii} is a linear combination of the s_k $(k \neq 0)$. Also, if $i \neq j$, b_{ij} is a linear combination of the s_k $(k \neq 0)$. It follows that

$$(3) \qquad | b_{ij} | = m_0{}^u s_0{}^n + B,$$

where u is a positive integer and B is a homogeneous polynomial in the s_k with integral coefficients, every term of which contains at least one s_k other than s_0.

Now let c_i be the co-factor of b_{i0} in $| b_{ij} |$, multiply the i-th equation (2) by c_i and add, thus showing that $| b_{ij} | = 0$. By the homomorphism h, it follows that $m_0{}^u \bar{s}_0{}^n + \bar{B} = 0$. But since $\bar{s}_k = 0$ $(k \neq 0)$, it is evident from the form of B that $\bar{B} = 0$. Thus $m_0{}^u \bar{s}_0{}^n = 0$, and hence $\bar{s}_0 = 0$.

We are now in a position to complete the proof of the lemma. By the first lemma, there exists an element ρ of C_k such that $g(\rho) = 0$ provided $g(x) \equiv 0(M)$. The correspondence

$$(4) \qquad \sum s_i a^i \rightarrow \sum \bar{s}_i \rho^i$$

then defines the required homomorphism $h' : S[a] \rightarrow D[\rho] = D_1$. For if $\sum s_i a^i = \sum r_i a^i$, then $\sum (\bar{s}_i - \bar{r}_i)x^i \equiv 0(M)$, $\sum \bar{s}_i \rho^i = \sum \bar{r}_i \rho^i$, and thus the correspondence (4) is actually independent of the manner in which an element of $S[a]$ is expressed as a polynomial in a with coefficients in S. It then follows readily that (4) defines a homomorphism, and in fact an extension of h as required.

In view of the theorem stated in the introduction, we shall establish Theorem 2 by showing the existence of a homomorphism of R into a subring of C_k taking any prescribed element $a \neq 0$ of R into a non-zero element of C_k.

Let $\{e, a\}$ denote the subring of R generated by a and e, the elements are therefore polynomials in a with coefficients in I. The set of polynomials $\sum n_i x^i$ in $I[x]$ such that $\sum n_i a^i = 0$ is an ideal N in $I[x]$ which clearly contains no non-zero element of I. We remark also that N contains no element of the form nx^m $(ne \neq 0)$. For, if the contrary were true, then $na^m = 0$ and since it is assumed that R has no nilpotent element it would follow that

$a = 0$. Thus there exists, by Lemma 1, an element $\sigma \neq 0$ of C_k such that if $g(x) \equiv 0(N)$, then $g(\sigma) = 0$. It now follows that the correspondence

$$\Sigma n_i a^i \rightarrow \Sigma n_i \sigma^i$$

is a homomorphism h of $\{e, a\}$ into $I[\sigma] \subset C_k$, by which $a \rightarrow \sigma \neq 0$. Except for showing that we may choose $\sigma \neq 0$, it will be noted that this is merely a special case of Lemma 2 in which S is the ring I.

It is now clear how to proceed. We assume that the elements of R are well ordered, with the first two elements as e and a respectively, where a is an arbitrary element of R other than zero. Let $\{e, a\}$ be the S of Lemma 2 and suppose b is the first element of R not in $\{e, a\}$. Then the homomorphism $h: \{e, a\} \rightarrow I[\sigma]$ can be extended to a homomorphism $\{e, a, b\} \rightarrow I[\sigma, \delta] \subset C_k$. We now may apply the lemma again with $\{e, a, b\}$ replacing the S. By transfinite induction it is readily shown that h can be extended to a homomorphism of R into a subring of C_k.[7] There thus exists a homomorphism of R into a subring of C_k which takes the arbitrary element $a \neq 0$ of R into an element $\sigma \neq 0$ of C_k. The proof of the theorem is therefore completed.

It will be noted that the ρ introduced in the proof of Lemma 2 is a root of the equation $f_a(x) = 0$, since $f_a(a) = 0$ and therefore $f_a(x) \equiv 0(M)$. Let us now assume that R is of characteristic p, and that *every* element of R satisfies the equation $x^{p^n} - x = 0$. It is readily verified that every element of the ring R' introduced in Theorem 1 also satisfies this same equation. By the above construction of a homomorphism of R into a subring of C_p, the only subrings of C_p which enter are those obtainable from $I = GF(p)$ by adjunction of roots of the equation $x^{p^n} - x = 0$. But the roots in C_p of this equation are precisely the elements of the Galois field $GF(p^n)$. We have thus established the following theorem.

THEOREM 3. *If the commutative ring R has characteristic p and every element satisfies the equation $x^{p^n} - x = 0$, then R is isomorphic to a subring of a direct sum of $GF(p^n)$.*

In the statement of this theorem it is not necessary to assume that R has no nilpotent elements, as it is readily verified that $a^m = 0$ is incompatible with $a^{p^n} = a$ unless $a = 0$.

This theorem will be seen to be a generalization of a result obtained in the joint paper with Montgomery, referred to in the introduction. In particular, if we choose $p = 2$, $n = 1$, this is equivalent to the theorem of Stone which states that a Boolean ring is isomorphic to an algebra of subclasses of some class, addition of classes being carried out modulo 2.

---

[7] See Stone, *loc. cit.*, p. 102.

5. Rings with all elements of finite order. Let S be an arbitrary ring all of whose elements are of finite order, and denote by S_p the set of elements of S whose orders are powers of the prime p. It follows easily that S_p is an ideal in S and that if $p \neq q$, S_p and S_q have no element except zero in common. From this it follows that the product of an element of S_p by one of S_q is always zero.

Let a be an arbitrary element of S, and $n = p_1^{\alpha_1} p_2^{\alpha_2} \cdots p_k^{\alpha_k}$ its order. Then the numbers $n/p_1^{\alpha_1}, \cdots, n/p_k^{\alpha_k}$ are relatively prime and there therefore exist integers $\beta_1, \cdots, \beta_k$ such that

$$\beta_1 n/p_1^{\alpha_1} + \cdots + \beta_k n/p_k^{\alpha_k} = 1.$$

From this it follows that if $a_{p_i} = (\beta_i n/p_i^{\alpha_i})a$, then a_{p_i} is an element of S_{p_i} and $a = \sum_{i=1}^{k} a_{p_i}$. Thus each element of S is expressible as a sum of a finite number of elements belonging respectively to S_p $(p = 2, 3, 5, \cdots)$. In fact it may be verified that each element can be so expressed in a unique way.

Now let P be the set of all primes p. In accordance with the concept of direct sum used above we shall define *the direct sum* of the rings S_p $(p = 2, 3, 5, \cdots)$ to be the ring of all functions defined on P, such that on p the values of the functions are in S_p. In this connection we have the following theorem.

THEOREM 4. *A ring S all of whose elements are of finite order is isomorphic to a subring of the direct sum of the ideals S_p $(p = 2, 3, 5, \cdots)$.*[8]

To establish this theorem we need only to exhibit an isomorphic ring of functions of the required type. If a is an arbitrary element of S we make correspond to it the function $f_a(p)$ defined as follows. If p is not a divisor of the order of a, then $f_a(p) = 0$, while if p is a divisor of the order of a, then $f_a(p) = a_p$ as defined above. The set of all functions f_a, where a ranges over the elements of S, is readily shown to be the required ring. For the discussion above shows that $f_{ab}(p) = f_a(p)f_b(p)$ and $f_{a+b}(p) = f_a(p) + f_b(p)$. That the correspondence $a \to f_a$ is actually an isomorphism follows from the fact that if $a \neq 0$ then $f_a(p) \neq 0$ for some p.

---

[8] It will be noted that S is in fact isomorphic to the ring of all functions defined on P, such that on p the functional values are in S_p, with the restriction that all functions differ from zero on only a finite number of points of P. From our point of view S is therefore a proper subring of the direct sum of the S_p although from other points of view S might well be called the direct sum of the S_p. See, e. g., Leo Zippin, "Countable torsion groups," *Annals of Mathematics*, vol. 36 (1935), pp. 86-99.

Let us now assume that S has no nilpotent elements. Then the order of no element can have a square factor. For if $k^2la = 0$, $kla \neq 0$, then $(kla)^2 = 0$, contrary to the assumption that there are no nilpotent elements. It follows that the ring S_p has characteristic p. From Theorems 2 and 4 we then obtain the following result.

THEOREM 5. *An algebraic ring R without nilpotent elements, all of whose elements are of finite order, is isomorphic to a subring of a direct sum of fields C_p, where p ranges over the primes.*

6. Ideal arithmetic in rings of characteristic p. We conclude with a brief study of ideal arithmetic in a ring of characteristic p. Let R be an algebraic ring of characteristic p with no nilpotent elements. If a is an arbitrary element of R, the subring $\{a\}$ generated by a is a finite commutative ring without nilpotent elements, and is therefore a direct sum of finite fields of characteristic p, say $GF(p^{n_i})$ $(i = 1, 2, \cdots, k)$.[9] If n is the L. C. M. of the n_i, then every element of $\{a\}$, in particular a itself, satisfies the equation $x^{p^n} - x = 0$.

Let N be an ideal in R, other than the unit ideal R itself, and consider the quotient ring R/N. This ring is a commutative ring of characteristic p, every element of which satisfies an equation $x^{p^n} - x = 0$, for proper choice of n. It follows that R/N contains no nilpotent elements and thus by Theorem 2, there is a homomorphism of R/N into a subring of C_p which takes any prescribed non-zero element of R/N into a non-zero element of C_p. We may now prove the following theorem.[10]

THEOREM 6. *Let R be an algebraic ring of characteristic p with no nilpotent elements. If M and N are ideals in R such that N is not a divisor of M, then there exists a prime ideal P in R which is a divisor of N but not of M.*

Let a be an element of R which is in M but not in N, and denote by $\bar{a}$ the image of a under the homomorphism $R \to R/N$. Thus $\bar{a} \neq 0$, and by the above discussion there exists a homomorphism $R/N \to C'_p$, where C'_p is a subring of C_p, taking $\bar{a}$ into an element of C_p different from zero. Thus the correspondence $R \to R/N \to C'_p$ defines a homomorphism of R into C'_p taking a into a non-zero element. This means, however, that there exists an ideal P in R such that $R/P \cong C'_p$. Since $C'_p \subset C_p$ has no divisors of zero, it follows that P is a prime ideal.

---

[9] van der Waerden, *op. cit.*, II, p. 163.

[10] Stone, *loc. cit.*, p. 105.

It is now easy to establish the final theorem.[10]

THEOREM 7. *Let R be an algebraic ring of characteristic p with no nilpotent elements. Then in R every ideal other than the unit ideal is the product of all its prime ideal divisors.*

Let $K \neq R$ be an ideal in R. Then K is not a divisor of R, and by the preceding theorem there exists a prime ideal containing K, so that K actually has prime ideal divisors. Let L be the product (intersection) of all the prime ideal divisors of K. Then clearly $K \subset L$. If L were not contained in K, there would be a prime ideal containing K but not L, contrary to the definition of L. Thus $L \subset K$, from which we conclude that $L = K$, and the theorem is proven.

SMITH COLLEGE.

METABELIAN GROUPS AND TRILINEAR FORMS.*

By Robert M. Thrall.

1. Introduction. H. R. Brahana has shown[1] that every metabelian group, G, composed of operators all except identity of prime order p, which contains H (an abelian group of order p and type $1, 1, 1, \cdots$) as a maximal invariant abelian subgroup and is generated by H and m independent permutable operators $u_1, \cdots, u_m$ from the group of isomorphisms of H, determines a trilinear form with coefficients in the $GF[p]$:

$$F(x, y, z) = \sum_{h,i,j} a_{hij} x_h y_i z_j,$$

and that conversely every such trilinear form determines a metabelian group having the above properties. We may suppose that G is generated by

$$H = \{s_1, \cdots, s_k, r_1, \cdots, r_l, t_1, \cdots, t_{n-k-l}\}$$

and $U = \{u_1, \cdots, u_m\}$ where $\{r_1, \cdots, r_l, t_1, \cdots, t_{n-k-l}\}$ is the central, $K = \{r_1, \cdots, r_l\}$ is the commutator subgroup, and $S = \{s_1, \cdots, s_k\}$ has no operators invariant under U. Then G is the direct product of $G' = \{S, U\}$ and $T = \{t_1, \cdots, t_{n-k-l}\}$. It is evident that classification of the groups G' implies the classification of all the groups G; so it is sufficient to consider only the groups G', and, henceforth we shall do so, (but will for simplicity in notation drop the primes).

The trilinear form $F(x, y, z)$ which has for a_{hij} the exponent of r_h in the commutator of s_j and u_i is completely determined by the choice of generators of K, H, and U. A different choice of generators would in general determine a different form F'. For instance, if new generators $u'_1, \cdots, u'_m$ of U are defined by the relations $u_i = u'^{\,a_{i1}}_1 \cdots \cdots u'^{\,a_{im}}_m$ $(i = 1, \cdots, m)$, F' would be obtained from F by applying the transformation

$$y'_i = a_{1i} y_1 + \cdots + a_{mi} y_m \qquad (i = 1, \cdots, m)$$

i. e. by applying the contragredient transformation on the y_i. Similarly changes of generators in S and in k would imply the contragredient transformations on the z_j and x_h respectively.

---

* Presented to the Society, April 9, 1937. Received by the Editors May 24, 1937.

[1] *Duke Mathematical Journal*, vol. 1 (1935), pp. 185-197. The results of this article will be used throughout the introduction without further reference.

Given a three-way matrix $M(a_{hij})$ there are six ways in which sets of variables x, y, z can be associated with the elements a_{hij} to give trilinear forms, these changes corresponding to interchanging the rôles played by the sets of variables x, y, z in the form $F(x, y, z)$. Each of these six interpretations of the matrix defines a group G. Of these six groups not more than three are abstractly distinct, since in G the subgroups S and U play abstractly the same rôles and hence can be interchanged without giving a new abstract group G. However, if l, m, k are all different at least three of the six groups G are distinct, since then there are commutator subgroups of three different orders: p^l, p^m, p^k.

2. Equivalence of trilinear forms and isomorphism of related groups G. We shall say that two forms $F_1(x, y, z)$ and $F_2(x, y, z)$ are *conjugate* or *equivalent* if, and only if, it is possible to transform F_1 into F_2 by means of linear transformations with coefficients in the $GF[p]$ on x, y, and z separately. A form $F(x, y, z)$ together with all the other forms conjugate to it will be said to constitute a *class of equivalent forms.* If F_1 and F_2 belong to the same class we write $F_1 \sim F_2$.

If G_1 is defined by $F_1(x, y, z)$ and G_2 is defined by $F_2(x, y, z)$, then G_1 is simply isomorphic to G_2 if either $F_2(x, y, z) \sim F_1(x, y, z)$ or

$$F_2(x, y, z) \sim F'_1(x, y, z) \equiv F_1(x, z, y)$$

where F'_1 has as its *h-matrices* [2] those of F_1 with the rows and columns interchanged. We define τ as the operation of interchanging y and z or in terms of the groups as the operation of interchanging S and U. If either $F_1 \sim F(x, y, z)$ or $F_1 \sim F'(x, y, z) \equiv F(x, z, y)$ we will write $F_1 \overset{\tau}{\sim} F$ and the totality of forms so related to a given one will be called a *τ-class.* It is evident that the groups corresponding to the forms of a τ-class are all simply isomorphic, and conversely if G_1 is simply isomorphic to G_2 and if the isomorphism can be exhibited by new choices of generators in S, U, and K and perhaps also use of τ, then the forms F_1 and F_2 belong to the same τ-class.

Two groups G_1 and G_2 are simply isomorphic if, and only if, generators $\{s'_1, \cdots, s'_k, u'_1, \cdots, u'_m\}$ of G_2 and $\{r'_1, \cdots, r'_l\}$ of K_2 can be chosen so that if $s_j^{-1}u_i^{-1}s_ju_i = \prod_{h=1}^{l} r_h^{a_h}$ in G_1 then $s'_j{}^{-1}u'_i{}^{-1}s'_ju'_i = \Pi r'_h{}^{a_h}$, i. e., if generators of G_2 and K_2 can be so chosen that the trilinear forms of G_1 and G_2 are the same. We associate with a given group G a collection E of forms, one form for each set of generators of G which satisfy our initial conditions as to maximal subgroups etc. Two groups are simply isomorphic if, and only if, the two collections which they define are the same.[3] From the preceding discussion it is clear that if one member of a τ-class D is in a collection E, then the whole τ-class is in E. If D coincides with E for a group G, then G is only isomorphic to other groups having the same τ-class. However, in general D does not coincide with E. For the numbers m and k previously defined, considered as an unordered pair, are invariants of a τ-class but are not necessarily invariants of the group G which defines the τ-class. For instance, the group $G = \{S, U\}$ of order p^{2+4+4} defined by $s_1^{-1}u_1^{-1}s_1u_1 = r_1$, $s_1^{-1}u_2^{-1}s_1u_2 = r_2$, $s_2^{-1}u_3^{-1}s_2u_3 = r_3$, $s_2^{-1}u_4^{-1}s_2u_4 = r_4$; the operators otherwise permutable, is likewise generated by the subgroups $S' = \{s_1, u_3, u_4\}$ and $U' = \{s_2, u_1, u_2\}$ which also satisfy the hypotheses of the first paragraph (§ 1). From the first definition of G we get $m = 2$, $k = 4$ whereas from the second we get $m = k = 3$. But in any case a collection E can be written as a sum of (distinct) τ-classes, so a determination of τ-classes is a first step toward the classification of groups G. We proceed along this line postponing the more general problem of classification of collections E.

---

[2] *Ibid.*, p. 190.

We first consider the changes induced in the matrix $M(a_{hij})$ when the variables x, y, and z undergo separate linear transformations. Let 1) $z_j = \Sigma\alpha_{j\nu}z'_\nu$ $(j = 1, 2, \cdots, k)$. Then

$$F(x, y, z) = \sum_{h,i,j} a_{hij}x_hy_iz_j = \sum_{h,i,\nu} a'_{hi\nu}x_hy_iz'_\nu$$

where

$$a'_{hi\nu} = \sum_j a_{hij}\alpha_{j\nu}.$$

Now consider the two dimensional matrix

$$M_x = (\sum_h a_{hij}x_h).$$

Under 1) this matrix is replaced by $M_x \cdot (\alpha_{j\nu})$. Similarly if 2) $y_i = \sum_\mu \beta_{\mu i}y'_\mu$ $(i = 1, 2, \cdots, m)$, F becomes $F' = \sum_{h,\mu,j} a'_{h\mu j}x_hy'_\mu z_j$ where $a'_{h\mu j} = \sum_i \beta_{\mu i}a_{hij}$ and M_x is replaced by $(\beta_{\mu i}) \cdot M_x$. Evidently M_x completely defines $M(a_{hij})$ and conversely. Under changes in y and z, M_x may be replaced by any matrix PM_xQ, where P and Q, of ranks m and k respectively, are any non-singular square matrices with elements in the $GF[p]$. Under τ the rows and columns of M_x are interchanged. If m and k are not equal, two forms (l, m, k) which belong to different classes will also belong to different τ-classes. Hence, in this case, we need only consider the classes (l, m, k) where, say, $m > k$ to obtain the τ-classes (l, m, k). $M_x = (\sum_h a_{hij}x_h) = \sum_h (a_{hij})x_h$. If we replace x by x' where $x_h = \sum_\lambda \gamma_{h\lambda}x'_\lambda$ we have $M_x = \sum_h (a_{hij}) \cdot \sum_\lambda \gamma_{h\lambda}x'_\lambda = \sum_\lambda (a'_{\lambda ij})x'_\lambda$ where $a'_{\lambda ij} = \sum_h \gamma_{h\lambda}a_{hij}$. That is, the h-sections of $M(a_{hij})$ are replaced by linear combinations of themselves under changes in x. The minors of M_x, polynomials in $x_1, \cdots, x_l$, are replaced by linear combinations of themselves under changes in y and z, and τ. And under changes in x these polynomials are replaced by polynomials conjugate to them under linear transformations on x. Hence, the projective invariants of the "invariant factors" of M_x will be invariants of the class or τ-class to which M_x belongs.

---

[3] Brahana's statement (*loc. cit.*, p. 189), concerning isomorphism of groups defined by two forms is incorrect and should be replaced by the above.

An analogous argument will generalize the results of the two preceding paragraphs to M_y and M_z, the arguments being identical in the case of classes but differing somewhat in the case of τ-classes.

We now generalize the concept of h-sections. We say that M_x for a fixed value a of x (i. e. $x_h = a_h$, $h = 1, 2, \cdots, l$) is an *x-section* of M_x and likewise of $M(a_{hij})$. We denote this by $M_x(a)$. The h-sections are then

$$M_x(1, 0, \cdots, 0), \cdots, M_x(0, 0, \cdots, 1).$$

We note that the rank of an x-section is unchanged under changes on y and z, and τ. But evidently the totality of x-sections is unchanged by linear transformation on x. Hence, the number of x-sections of any given rank is an invariant of M_x and also of $M(a_{hij})$. Similarly the numbers of y-sections or z-sections of any given rank are invariants of $M(a_{hij})$, and of the τ-classes aside from the possibility of interchanging the numbers for y and z under τ.

If $m = k$, $F(x, y, z)$ and $F(x, z, y)$ might be equivalent. Hence, we find the classes with invariants l, m, m, and then combine such of those as are equivalent under τ. Furthermore, for $m = k$ the matrix M_x is square, and its determinant, the only m-rowed minor, is a form $f(x)$ of degree m in the l variables x. The projective invariants of $f(x)$ will be invariants of the τ-class to which M_x belongs. This gives particular interest to the forms and groups having $m = k$, and the major portion of this discussion will be devoted to the case $m = k = 3$.

We propose to classify the groups G and related forms F by classifying the matrices M_x under multiplication on left and right by non-singular matrices with constant elements, linear transformations on x, and τ if $m = k$. This scheme will give each abstract group G once for each τ-class that it defines, and hence each group at least once. In the case $m = k = 3$ we shall complete the classification by finding the groups that can be defined by more than one τ-class.

That a matrix M_x should correspond to a group G imposes certain restrictions on it. First, there must be one element in each row of M_x different from zero, since no u_i is permutable with every s_j. Similarly there must be one element in each column different from zero. This must be true not only for M_x but also for every M'_x in the same τ-class as M_x. (We say that M_x and M'_x are in the same τ-class if the forms corresponding to them belong to the same τ-class).

Brahana[4] has shown that no one of the numbers l, m, k can be greater than the product of the other two, and that there is just one group when one of the numbers is the product of the other two.

3. Apolarity of trilinear forms. Let us write $r_{ij} = u_i^{-1} s_j^{-1} u_i s_j$. The mk commutators r_{ij} generate K, of order p^l, and hence satisfy $mk - l$ independent relations which can be expressed in the form $\prod_{i,j} r_{i,j}^{\alpha_{\nu ij}} = 1$, $\nu = 1, \cdots, mk - l$, or replacing the subscript i, j by $\lambda = (i - 1)k + j$ this becomes $\prod_{\lambda} r_\lambda^{\alpha_{\nu\lambda}} = 1$, $\nu = 1, \cdots, mk - l$. For these relations to be independent the matrix $(\alpha_{\nu\lambda})$ must be of rank $mk - l$. In the group G^* with $l = mk$ the commutators r_{ij} satisfy no relations and can always be taken as a set of independent generators for K. That is, in this group the expression for any operator of K in terms of the r_{ij} is unique.

We shall show that *any group G with K of order p^l is a quotient group of the group G^* with $l = mk$, where p^m and p^k are the orders of U and S in G and G^*.* Consider the quotient group of G^* with respect to an operator $\bar{r} = \prod_{\lambda} r_\lambda^{\beta_\lambda}$. ($\bar{r}$ is in the central since G^* is metabelian). This group $G^*/\bar{r}$ has $l = mk - 1$ and the relation satisfied by its commutators is $\prod_{\lambda} r_\lambda^{\beta_\lambda} = 1$. The quotient group G^*/R where $R = \{\bar{r}_1, \cdots, \bar{r}_{mk-l}\}$, $\bar{r}_\nu = \prod_{\lambda} r_\lambda^{\beta_{\nu\lambda}}$, $\nu = 1, \cdots, mk - l$ has its relations given by the rows of the matrix $(\beta_{\nu\lambda})$. As the $\beta_{\nu\lambda}$ are at our disposal they may be chosen as the $\alpha_{\nu\lambda}$ of the relations connecting the commutators r_λ of the given group G. And as two groups whose commutators $r_\lambda \equiv r_{ij}$ satisfy the same relations are certainly simply isomorphic, the theorem follows. Further, the quotient group G^*/R can be obtained by taking successive quotient groups of index p, i. e. first $G_1 = G^*/\bar{r}_1$ then $G_2 = G_1/\bar{r}_{12}$ and so on until we reach $G_{mk-l} = G^*/R$ (where $\bar{r}_{12}$ in G_1 corresponds to $\bar{r}_2$ in G^* etc.). *Hence any group G with K of order p^l is a quotient*

---

[4] *Ibid.*, p. 191.

group of at least one group G with K of order p^{l+1} (provided $l < mk$). These successive quotient groups could just as well have been taken with respect to any set of generators of R (and the operators corresponding to them in the successive quotient groups). If these new generators $\bar{r}'_h$ are given in terms of the $\bar{r}_\nu$ by the relations $\bar{r}'_h = \prod_\nu \bar{r}_\nu^{\gamma_{h\nu}}$ and hence in terms of the r_λ by the relations

$$\bar{r}'_h = \prod_\nu \left(\prod_\lambda r_\lambda^{\alpha_{\nu\lambda}} \right)^{\gamma_{h\nu}} = \prod_\lambda r_\lambda^{\alpha'_{h\lambda}}$$

where $(\alpha'_{h\lambda}) = (\gamma_{h\nu})(\alpha_{\nu\lambda})$, the matrix $(\alpha'_{h\lambda})$ defines the same group G as $(\alpha_{\nu\lambda})$. Further, if we let $\alpha_{\nu\lambda} = \alpha_{\nu ij}$ and $\alpha'_{h\lambda} = \alpha'_{hij}$, where as before $\lambda = (i-1)k + j$, then the forms

$$F(x, y, z) = \Sigma \alpha_{\nu ij} x_\nu y_i z_j \quad \text{and} \quad F' = \Sigma \alpha'_{hij} x_h y_i z_j$$

belong to the same τ-class. Choice of a new set of generators in U and S corresponds as before to linear transformations on y and z; so if a form F describes the relations satisfied by the commutators r_{ij} then any other form $F' \overset{\tau}{\sim} F$ will describe the relations of a group G' simply isomorphic with G.

Thus we have the group G defined by two sets of forms: the one giving the relations satisfied by the commutators $r_\lambda \equiv r_{ij}$ and the other, the original one, giving the expression for the r_{ij} in terms of any given set of generators of K. We shall establish certain "apolarity" relations between these two sets of forms (or rather between their matrices) which will cut in half the work necessary for determining the τ-classes, and hence, the groups G, with numbers l, m, k.

Suppose that a given group G is defined in the two above ways by the relations 1) $\prod_{i,j} r_{\substack{b_{\nu ij} \\ ij}} = 1$, $\nu = 1, \cdots, mk - l$, and 2) $r_{ij} = \prod_{h=1}^{l} \bar{r}_h^{a_{hij}}$, $i = 1, \cdots, m$; $j = 1, \cdots, k$. Substituting 2) in 1) we have

$$\prod_{i,j} \left(\prod_{h=1}^{l} \bar{r}_h^{a_{hij}} \right)^{b_{\nu ij}} = \prod_{h=1}^{l} \bar{r}_h^{\sum\limits_{i,j} a_{hij} b_{\nu ij}} = 1.$$

But since the $\bar{r}_h$ are independent this requires 3) $\sum_{i,j} a_{hij} b_{\nu ij} = 0$, $h = 1, \cdots, l$; $\nu = 1, \cdots, mk - l$. Two matrices satisfying 3) will be called *apolar*; likewise the two τ-classes defined by them will be called *apolar*.

Since the apolarity condition is symmetric in the two matrices (a_{hij}) and $(b_{\nu ij})$ we may by interchanging their rôles obtain a new group G' of order $p^{mk-l+m+k}$ which group we will say is *apolar* to the initial group G.[5] It is evident that the classification of the groups $G(l, m, k)$ implies the classification of the groups $G(mk - l, m, k)$ and hence that for purposes of classification we never need consider l larger than $mk/2$.

---

[5] If neither a given τ-class nor its apolar τ-class implies that one of $\{S, K\}$ and

4. General theory for the groups and forms $(l, 3, 3)$. We shall summarize the preceding argument by applying it to the case $m = k = 3$, $l = 0, 1, \cdots, 9$. The matrix M_x becomes $M_x = (\sum_{h=1}^{l} a_{hij}x_h)$ where we need consider only $l = 0, 1, 2, 3, 4$. Under changes of generators in U and in S, M_x is replaced by PM_xQ where P and Q are non-singular constant three rowed matrices. Under τ, M_x is replaced by its transposed. Under changes on x the element a_{ij} of M_x is replaced by $a'_{ij} = \sum_{\lambda} a'_{\lambda ij}x'_{\lambda}$ where $x_h = \sum_{\lambda} \alpha_{\lambda h}x'_{\lambda}$ and $a'_{\lambda ij} = \sum_{h} \alpha_{\lambda h}a_{hij}$. Further, $| M_x | = f(x)$, a cubic form in $x_1, \cdots, x_l$, is replaced by $f_1(x')$, a form conjugate to $f(x)$ under the given transformation. So, first, we classify cubic forms in l variables. Then choosing a particular form from each projective class we ask how many τ-classes can have this cubic for $| M_x |$. Given two matrices M_x and M'_x with $| M_x | = | M'_x | = f(x)$ the forms F and F' are equivalent if $PM_xQ = M'_x$. If $f(x) \equiv g(x') = cf(x')$ we say that the transformation from x to x' is an *automorphism* of $f(x)$. The totality of such transformations constitutes a group A, called the *group of automorphisms* of $f(x)$. M_x and M'_x are equivalent if, and only if, $PM''_xQ = M'_x$ where M_x is replaced by M''_x under some automorphism of $f(x)$. $M_x \overset{\tau}{\sim} M'_x$ if either $M_x \sim M'_x$ or $M_x \sim M'_x$-transpose. (In what follows we shall be concerned chiefly with M_x and will drop the subscript, using $M(x_1, \cdots, x_l)$ instead of $M_x(x_1, \cdots, x_l)$. We shall continue to use the subscripts in $M_y(y)$ and $M_z(z)$.)

5. $l = 0$. Classification of commutators. For $l = 0$ the only τ-class is the null class. It defines the abelian group $\{S, U\}$ of order p^6 which is not a group G, (see footnote 5), and the group $G^* = \{S, U\}$ of order p^{6+9} which is as we have seen the only group G with the numbers 9, 3, 3. As the other groups $G(l, 3, 3)$ are all quotient groups of this group G^* with respect to subgroups of K an analysis of the nature of the operators in G^* will throw light upon the group theoretic meaning of the invariants of the τ-classes. The succeeding arguments given for the case $(l, 3, 3)$ are capable of direct generalization to the general case (l, m, k).

$\{U, K\}$ shall not be maximal abelian, the two groups defined by it are both groups G. Otherwise it may happen that one (or both) of these groups will not be in the set of groups G, although it will be the direct product of a group G and an abelian group of order p^{α} and type $1, 1, 1 \cdots$.

In a group G we say that the commutator of u^y (i. e. $u_1^{y_1}u_2^{y_2}u_3^{y_3}$) and s^z is a *commutator of the first kind.*[6] An operator in K which is not a commutator of the first kind but which is a product of two commutators of the first kind will be called a *commutator of the second kind.* Now any operator, T, in G can be written as $u^y s^z r^x$. The commutator of T and any other operator $T' = u^{y'} s^{z'} r^{x'}$, $T^{-1}T'^{-1}TT'$, is

$$\begin{aligned} r^{-x}s^{-z}u^{-y}\cdot r^{-x'}s^{-z'}u^{-y'}\cdot u^y s^z r^x\cdot u^{y'}s^{z'}r^{x'} &= s^{-z}(u^{-y}s^{-z'}u^y s^{z'})s^{-z'}(u^{-y'}s^z u^{y'}s^{-z})s^z s^{z'} \\ &= (u^{-y}s^{-z'}u^y s^{z'})(u^{-y'}s^z u^{y'}s^{-z}). \end{aligned}$$

Hence, *every commutator in a group G is of kind* 0, 1, *or* 2. In general these do not include all of the operators of K, and we shall say that an operator of K is a *commutator of the ν-th kind* if it is a product of ν but of no less than ν commutators of the first kind. (In a group $G(l, m, k)$ no commutator is of kind greater than the smaller of m and k.)

The general commutator of the first kind in the group $G^*(9, 3, 3)$, that of u^y and s^z, has as its matrix[7]

$$\begin{pmatrix} y_1z_1 & y_1z_2 & y_1z_3 \\ y_2z_1 & y_2z_2 & y_2z_3 \\ y_3z_1 & y_3z_2 & y_3z_3 \end{pmatrix} = \begin{pmatrix} y_1 & 0 & 0 \\ 0 & y_2 & 0 \\ 0 & 0 & y_3 \end{pmatrix}\begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix}\begin{pmatrix} z_1 & 0 & 0 \\ 0 & z_2 & 0 \\ 0 & 0 & z_3 \end{pmatrix}$$

and hence is of rank one provided that neither u^y nor s^z is identity. Conversely, *any operator in K with a matrix (α_{ij}) of rank one is a commutator of the first kind,* since, then, there exist numbers y_1, y_2, y_3; z_1, z_2, z_3 such that $\alpha_{ij} = y_i \cdot z_j$.

If r^α is a commutator of the second kind in G^*, its matrix (α_{ij}) will be the sum of the matrices for the commutators of the first kind of which it is the product. If (α_{ij}) were of rank less than two, r^α would be of kind less than two, and since the sum of two matrices of ranks k_1 and k_2 is of rank $k_1 + k_2$ at most, we have that *the matrix of a commutator of the second kind in G^* is of rank two.* Conversely, *an operator in K with matrix of rank two is of the second kind.* For if (α_{ij}) is of rank two there exist non-singular matrices P and Q such that $(\alpha_{ij}) = P\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix}Q$, whence

$$(\alpha_{ij}) = P\begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}Q + P\begin{pmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix}Q$$

---

[6] Identity is said to be of the zero-th kind.

[7] The element in the i-th row and j-th column is the exponent of r_{ij} in this commutator. Every such matrix defines an element of K.

is an expression of (α_{ij}) as the sum of two matrices of rank one, giving r^a as the product of two commutators of the first kind. Similarly, *a necessary and sufficient condition that an operator in K shall be a commutator of the third kind is that its matrix be of rank three.*[8] In G^* we have therefore as many non-commutators in the commutator subgroup K as there are three rowed matrices of rank three. The ratio of this number to the order of K, approaches the limit one as the prime, p, becomes infinite.[9]

Now consider a commutator $r^a = \prod_{i,j} r_{ij}^{a_{ij}}$ in a group $G(mk - l, m, k)$. For $l > 0$ the matrix (a_{ij}) is not uniquely defined by r^a. For identity in G corresponds to the invariant subgroup $R = \{\prod_{i,j} r_{ij}^{a_{hij}}\}$, $h = 1, \cdots, l$, in $G^*(mk, m, k)$ and if r^a corresponds to a particular commutator $r'^{a'}$ in G^* then it also corresponds to those and only to those commutators in the coset $Rr'^{a'}$, each operator of which defines one of the p^l matrices (a_{ij}) which will represent r^a. If (a_{ij}) is any particular matrix which represents r^a the others which represent it are of the form $(a_{ij}) + M(x)$, x arbitrary.

If r^a is a commutator of the ν-th kind then it must correspond to at least one commutator of the ν-kind in G^* (and to none of kind less than ν) and therefore can be represented by at least one matrix (a_{ij}) of rank ν. Now suppose that the maximum rank of the matrices representing r^a is $\nu + \nu'$. Then we shall say that *r^a is a commutator of the ν-th kind and extent ν'.*

Two groups G_1 and G_2 are isomorphic if and only if R_1 and R_2 are conjugate under some isomorphism of G^*. But in any isomorphism of G^* with itself, a commutator of the ν-th kind corresponds to a commutator of the ν-th kind, for the rank of its matrix is unchanged by new choice of generators in S, U, and K. So the number of commutators of the ν-th kind in G^* corresponding to identity in G is an invariant of G.

If two commutators r_1, r_2 in G give rise to non-isomorphic quotient groups G/r_1 and G/r_2, we are justified in distinguishing between r_1 and r_2 in G. The following theorem gives such justification to the definition of kind and extent.

THEOREM. *If in a group G two commutators*[10] *r_1 and r_2 are such that G/r_1 and G/r_2 are isomorphic, then r_1 and r_2 are of the same kind and extent.*

---

[8] An obvious generalization to all groups G with $l = mk$ is: *A necessary and sufficient condition that an operator in the commutator subgroup be of the ν-th kind is that its matrix shall be of rank ν.*

[9] It is of some interest to contrast this result with early conjectures with respect to the existence of non-commutators in the commutator subgroup. See W. B. Fite, *Transactions of the American Mathematical Society*, vol. 3 (1902), pp. 331-353, in particular pp. 332 and 339.

[10] For simplicity we shall call any operator in K a commutator, recalling, of course, that it is not a commutator in the ordinary sense if its kind is greater than two.

Let r_i be of kind ν_i and extent ν'_i. Let identity in G/r_i correspond to $\{R, r'_i\}$ in G^*, where r'_i is of kind ν_i and $G^*/R = G$. Then if $\nu_1 \neq \nu_2$, say $\nu_1 < \nu_2$, $\{R, r'\}$ has more operators of kind ν_1 than $\{R, r'_2\}$, for as a consequence of the definition of kind no operator in $\{R, r'_2\}$ can be of kind less than ν_2 unless it is in R. Hence $\nu_1 = \nu_2$. Next suppose $\nu_1 = \nu_2$, but $\nu'_1 \neq \nu'_2$, say $\nu'_1 < \nu'_2$. Then $\{R, r'_2\}$ contains an operator of kind $\nu_1 + \nu'_2$ whereas no operator in $\{R, r'_1\}$ is of kind greater than $\nu_1 + \nu'_1$. Hence, the isomorphism of G/r_1 and G/r_2 implies $\nu_1 = \nu_2$ and $\nu'_1 = \nu'_2$.

The converse of this theorem is not true as is illustrated by the groups represented by the forms $M_{10}(x)$ and $M_9(x)$ (§ 7). These are both quotient groups of M_2 (§ 6) with respect to commutators of the second kind and extent zero.

6. τ-classes (1, 3, 3) and related groups.

For $l = 1$, $M(x) = (a_{1ij}x_1)$ and the τ-class is completely defined by the rank of (a_{1ij}); so we have three τ-classes, represented by $M_1 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$, $M_2 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix}$ and $M_3 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$.

Interpreted for $l = 1$, the first two do not define groups G. M_3 does give a group G, the only one for 1, 3, 3. Interpreted for $l = 8$ we get three groups G, say G_1, G_2, G_3, where $G_i = G^*/r^{x^{(i)}}$, $r^{x^{(i)}}$ being a commutator of the i-th kind in G^*, which gives a differentiation of the groups G according to the preceding theorem.

7. τ-classes (2, 3, 3) and related groups.

For $l = 2$, $M(x) = x_1(\alpha_{ij}) + x_2(\beta_{ij})$ and a complete classification is given by the theories of invariant factors and of binary cubics. For $M(x)$ we may have: 1) an irreducible cubic, 2) a cubic with one real root, 3) a cubic with three distinct real roots, 4) a cubic with two roots, one repeated, 5) a cubic with a triple root, 6) a cubic identically zero.

It has been shown[11] that all cubics $f(x)$ belonging to and irreducible in the $GF[p]$ are conjugate under the group of linear transformations on x_1 and x_2. Hence 1) gives just one τ-class represented by $M_1(x) = \begin{pmatrix} x_1 & x_2 & 0 \\ 0 & x_1 & x_2 \\ \alpha x_2 & \alpha x_2 & x_1 \end{pmatrix}$ where $x_1^3 - \alpha x_1 x_2^2 + \alpha x_2^3$ is irreducible. For 2) it is evident that we may take $f(x) = x_1(x_1^2 - \gamma x_2^2)$, $\gamma \neq$ square, giving the single τ-class represented by $M'_2(x) = \begin{pmatrix} x_1 & x_2 & 0 \\ 0 & x_1 & x_2 \\ 0 & \gamma x_2 & x_1 \end{pmatrix}$, or if we prefer by $M_2(x) = \begin{pmatrix} x_1 & 0 & 0 \\ 0 & x_1 & x_2 \\ 0 & \gamma x_2 & x_1 \end{pmatrix}$. For 3) we may take $f(x) = x_1x_2(x_1 + x_2)$ giving the τ-class represented by $M_3(x) = \begin{pmatrix} x_1 & 0 & 0 \\ 0 & x_2 & 0 \\ 0 & 0 & x_1 + x_2 \end{pmatrix}$. For 4) we may take $f(x) = x_1^2x_2$ and have two τ-classes distinguished by their invariant factors, represented by $M_4(x) = \begin{pmatrix} x_1 & 0 & 0 \\ 0 & x_1 & 0 \\ 0 & 0 & x_2 \end{pmatrix}$ and $M_5(x) = \begin{pmatrix} x_1 & x_2 & 0 \\ 0 & x_1 & 0 \\ 0 & 0 & x_2 \end{pmatrix}$. For 5) we take $f(x) = x_1^3$ and have (since x_2 must appear) the two τ-classes represented by $M_6(x) = \begin{pmatrix} x_1 & x_2 & 0 \\ 0 & x_1 & 0 \\ 0 & 0 & x_1 \end{pmatrix}$ and $M_7(x) = \begin{pmatrix} x_1 & x_2 & 0 \\ 0 & x_1 & x_2 \\ 0 & 0 & x_1 \end{pmatrix}$. (If x_2 did not appear we should have the case 1, 3, 3 instead of 2, 3, 3).

---

[11] H. R. Brahana, *Bulletin of the American Mathematical Society*, vol. 39 (1933), pp. 962-969.

For 6) $f(x) \equiv 0$ we list the following representatives of τ-classes with descriptive invariants sufficient to establish their distinctness:

$$M_8(x) = \begin{pmatrix} x_1 & x_2 & 0 \\ \gamma x_2 & x_1 & 0 \\ 0 & 0 & 0 \end{pmatrix},\ \gamma \neq \text{square}; \qquad M_9(x) = \begin{pmatrix} x_1 & x_2 & 0 \\ 0 & x_1 & x_2 \\ 0 & 0 & 0 \end{pmatrix};$$

$$M_{10}(x) = \begin{pmatrix} x_1 & 0 & 0 \\ 0 & x_1 & x_2 \\ x_2 & 0 & 0 \end{pmatrix}; \qquad M_{11}(x) = \begin{pmatrix} x_1 & x_2 & 0 \\ 0 & x_1 & 0 \\ 0 & 0 & 0 \end{pmatrix};$$

$$M_{12}(x) = \begin{pmatrix} x_1 & 0 & 0 \\ 0 & x_1 & x_2 \\ 0 & 0 & 0 \end{pmatrix}; \qquad M_{13}(x) = \begin{pmatrix} x_1 & 0 & 0 \\ 0 & x_2 & 0 \\ 0 & 0 & 0 \end{pmatrix};$$

$$M_{14}(x) = \begin{pmatrix} x_1 & x_2 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

Every x-section of M_{14} is of rank one, which is sufficient to distinguish it from the others. M_{13} has two x-sections of rank one; M_{12} and M_{11} each one; and M_{10}, M_9, M_8 no x-sections of rank one. M_{11} and M_{12} differ in that F_{11} is free of both y_3 and z_3, whereas F_{12} is free of y_3 but not of z_3. Similarly F_8 is free of y_3 and z_3; F_9 is free of y_3; and F_{10} involves all of the variables. The completeness of this list of τ-classes for $f(x) \equiv 0$ follows from Brahana's analysis of the cases: (3, 2, 3), (2, 2, 3), (2, 2, 2), (2, 2, 1).[12]

---

[12] *American Journal of Mathematics*, vol. 56 (1934), pp. 490-510. He omitted the group corresponding to M_8 and listed one non-existent group $G(2, 2, 3)$.

Interpreted for $l = 2$, $M_1, \cdots, M_7$ and M_{10} give the eight groups G with numbers 2, 3, 3 and the other forms do not give groups G. Interpreted for $l = 7$ we get the fourteen groups G with numbers 7, 3, 3.

8. $l = 3$; theory and method of attack. For $l = 3$ in addition to the ternary cubic $|M(x)|$ we have also the ternary cubics $|M_y|$ and $|M_z|$ (where M_y and M_z are defined by the y- and z-sections of $M(a_{hij})$ just as $M(x) \equiv M_x$ is defined by the x-sections).

First, we shall classify the forms according to the projective invariants of $f(x_1, x_2, x_3) = |M(x)|$. A. D. Campbell has[13] given a projective classification of ternary cubics with coefficients in the $GF[p]$. To determine the representations of a given cubic $f(x)$ as $|M(x)|$ we shall first represent a line section of it, say $f(x_1, 0, x_3)$, as a determinant according to the methods used for the case $l = 2$, giving $M(x_1, 0, x_3)$ such that $|M(x_1, 0, x_3)| = f(x_1, 0, x_3)$. Then we shall consider $M(x) = M(x_1, 0, x_3) + (a_{ij})x_2$ where the a_{ij} are to be determined so that $|M(x)| = f(x)$. For $f(x)$ we take in turn a representative from each projective class or set of classes of cubics (making several additions and corrections to Campbell's list). Aside from class 20 and the class $f(x) \equiv 0$, $f(x_1, 0, x_3)$ is one of the following binary cubics: $f_1 = x_1^3$, $f_2 = x_1^2(x_1 + x_3)$, $f_3 = x_1^2 x_3$, $f_4 = x_1(x_1^2 - x_3^2)$; and so we may suppose that $M_1(x_1, 0, x_3)$ is one of the following seven matrices:

$$M_1 = \begin{pmatrix} x_1 & x_3 & 0 \\ 0 & x_1 & 0 \\ 0 & 0 & x_1 \end{pmatrix}, \qquad M_2 = \begin{pmatrix} x_1 & x_3 & 0 \\ 0 & x_1 & x_3 \\ 0 & 0 & x_1 \end{pmatrix} \text{ from } f_1;$$

$$M_3 = \begin{pmatrix} x_1 & 0 & 0 \\ 0 & x_1 & 0 \\ 0 & 0 & x_1 + x_3 \end{pmatrix}, \qquad M_4 = \begin{pmatrix} x_1 & x_3 & 0 \\ 0 & x_1 & 0 \\ 0 & 0 & x_1 + x_3 \end{pmatrix} \text{ from } f_2;$$

$$M_5 = \begin{pmatrix} x_1 & 0 & 0 \\ 0 & x_1 & 0 \\ 0 & 0 & x_3 \end{pmatrix}, \qquad M_6 = \begin{pmatrix} x_1 & x_3 & 0 \\ 0 & x_1 & 0 \\ 0 & 0 & x_3 \end{pmatrix} \text{ from } f_3;$$

$$M_7 = \begin{pmatrix} x_1 & 0 & 0 \\ 0 & x_1 + x_3 & 0 \\ 0 & 0 & x_3 \end{pmatrix} \text{ from } f_4.$$

We now give the expansions of $|M_i(x_1, 0, x_3) + (a_{ij})x_2|$, $i = 1, \cdots, 7$, and determine the conditions on (a_{ij}) for

$$|M(x)| = f(x) = a_1 x_1^3 + a_2 x_2^3 + a_3 x_3^3 + a_4 x_1^2 x_2 + a_5 x_1 x_2^2 + a_6 x_1^2 x_3 + a_7 x_1 x_3^2 + a_8 x_2^2 x_3 + a_9 x_2 x_3^2 + a_{10} x_1 x_2 x_3,$$

---

[13] *Messenger of Mathematics*, vol. 58 (1928), pp. 33-48.

where the a_ν are given.

$|M_1(x_1, 0, x_3) + (a_{ij})x_2| = f(x)$, $a_1 = 1$, $a_3 = a_6 = a_7 = a_9 = 0$, gives B_1:

1) $a_{11} + a_{22} + a_{33} = a_4$ 3) $|a_{ij}| = a_2$

2) $a_{11}a_{22} + a_{11}a_{33} + a_{22}a_{33} - a_{12}a_{21} - a_{13}a_{31} - a_{23}a_{32} = a_5$

4) $a_{23}a_{31} - a_{21}a_{33} = a_8$ 5) $-a_{21} = a_{10}$

$|M_2(x_1, 0, x_3) + (a_{ij})x_2| = f(x)$, $a_1 = 1$, $a_3 = a_6 = a_7 = 0$, gives B_2: 1), 2), 3) same as in B_1.

4) $-a_{21}a_{33} + a_{23}a_{31} + a_{12}a_{31} - a_{11}a_{32} = a_8$

5) $a_{31} = a_9$ 6) $-a_{31} - a_{32} = a_{10}$

$|M_3(x_1, 0, x_3) + (a_{ij})x_2| = f(x)$, $a_1 = a_6 = 1$, $a_3 = a_7 = a_9 = 0$, gives B_3: 1), 2), 3) same as in B_1.

4) $a_{11}a_{22} - a_{21}a_{12} = a_8$ 5) $a_{11} + a_{22} = a_{10}$

$|M_4(x_1, 0, x_3) + (a_{ij})x_2| = f(x)$, $a_1 = a_6 = 1$, $a_3 = a_7 = 0$, gives B_4: 1), 2), 3) same as in B_1.

4) $a_{11}a_{22} - a_{12}a_{21} + a_{23}a_{31} - a_{21}a_{33} = a_8$

5) $-a_{21} = a_9$ 6) $a_{11} + a_{22} - a_{21} = a_{10}$

$|M_5(x_1, 0, x_3) + (a_{ij})x_2| = f(x)$, $a_6 = 1$, $a_1 = a_3 = a_7 = a_9 = 0$, gives B_5:

1) $a_{33} = a_4$ 3) $|a_{ij}| = a_2$

2) $a_{11}a_{33} + a_{22}a_{33} - a_{23}a_{32} - a_{13}a_{31} = a_5$

4) $a_{11}a_{22} - a_{12}a_{21} = a_8$ 5) $a_{11} + a_{22} = a_{10}$

$|M_6(x_1, 0, x_3) + (a_{ij})x_2| = f(x)$, $a_6 = 1$, $a_1 = a_3 = a_7 = 0$, gives B_6: 1), 2), 3) same as in B_5.

4) $a_{11}a_{22} + a_{23}a_{31} - a_{12}a_{21} - a_{21}a_{33} = a_8$

5) $-a_{11} = a_9$ 6) $a_{33} - a_{22} = a_{10}$

$|M_7(x_1, 0, x_3) + (a_{ij})x_2| = f(x)$, $a_1 = a_7 = 1$, $a_3 = a_6 = 0$, gives B_7: 1), 2), 3), same as in B_1.

4) $-a_{11}a_{22} + a_{11}a_{33} + a_{12}a_{21} - a_{13}a_{31} = a_8$

5) $-a_{11} = a_9$ 6) $a_{33} - a_{22} = a_{10}$.

Two solutions $M(x)$ and $M'(x)$ of $|M(x)| = f(x)$, such that $M(x_1, 0\ x_3) = M'(x_1, 0, x_3)$, belong to the same τ-class if $M' = PMQ$, which implies $PM(x_1, 0, x_3)Q = M(x_1, 0, x_3)$. We must therefore determine the pairs of matrices P and Q such that $PM_i(x_1, 0, x_3)Q = M_i(x_1, 0, x_3)$ for $i = 1, \cdots, 7$. To show that $PM_i(x_1, 0, x_3)Q = M_i(x_1, 0, x_3)$ identically in x_1 and x_3, it is sufficient to show that it is true for any independent pair of number couples $(x_1, 0, x_3)$. Proceeding thus we find that in each case $Q = P^{-1}$. Let P_i be the most general constant non-singular matrix permutable with M_i. Then a simple computation (which we shall omit) gives:

$$P_1=\begin{pmatrix}\alpha_{11} & \alpha_{12} & \alpha_{13}\\ 0 & \alpha_{11} & 0\\ 0 & \alpha_{32} & \alpha_{33}\end{pmatrix};\quad P_2=\begin{pmatrix}\alpha_{11} & \alpha_{12} & \alpha_{13}\\ 0 & \alpha_{11} & \alpha_{12}\\ 0 & 0 & \alpha_{11}\end{pmatrix};\quad P_3=\begin{pmatrix}\alpha_{11} & \alpha_{12} & 0\\ \alpha_{21} & \alpha_{22} & 0\\ 0 & 0 & \alpha_{33}\end{pmatrix};$$

$$P_4=\begin{pmatrix}\alpha_{11} & \alpha_{12} & 0\\ 0 & \alpha_{11} & 0\\ 0 & 0 & \alpha_{33}\end{pmatrix};\quad P_5=\begin{pmatrix}\alpha_{11} & \alpha_{12} & 0\\ \alpha_{21} & \alpha_{22} & 0\\ 0 & 0 & \alpha_{33}\end{pmatrix};\quad P_6=\begin{pmatrix}\alpha_{11} & \alpha_{12} & 0\\ 0 & \alpha_{11} & 0\\ 0 & 0 & \alpha_{33}\end{pmatrix};$$

$$P_7=\begin{pmatrix}\alpha_{11} & 0 & 0\\ 0 & \alpha_{22} & 0\\ 0 & 0 & \alpha_{33}\end{pmatrix};$$

the elements in each matrix being arbitrary, except for the negative restriction that none of the matrices be singular.

The matrices P_1 form a group (under matrix multiplication) generated by

$$T_{12}(m)=\begin{pmatrix}1 & m & 0\\ 0 & 1 & 0\\ 0 & 0 & 1\end{pmatrix},\quad T_{13}(m)=\begin{pmatrix}1 & 0 & m\\ 0 & 1 & 0\\ 0 & 0 & 1\end{pmatrix},\quad T_{32}(m)=\begin{pmatrix}1 & 0 & 0\\ 0 & 1 & 0\\ 0 & m & 1\end{pmatrix},$$

$$T_{33}(m)=\begin{pmatrix}1 & 0 & 0\\ 0 & 1 & 0\\ 0 & 0 & m\end{pmatrix},\quad T(m)=\begin{pmatrix}m & 0 & 0\\ 0 & m & 0\\ 0 & 0 & m\end{pmatrix},$$

m arbitrary in each generator. Transformation of a matrix (α_{ij}) by $T_{ij}(m)$ effects 1) adding m times the i-th column to the j-th column and 2) adding $-m$ times the j-th row to the i-th row. Let $t_{ij}(m)$ represent this operation of transformation. Transformation of (α_{ij}) by $T_{ii}(1/m)$ effects: 1) multiplying the i-th row by m and 2) multiplying the i-th column by $1/m$. Let $t_{ii}(m)$ represent this operation of transformation. Transformation by $T(m)$ effects no change. Then the group $T_1=\{t_{12}(m), t_{13}(m), t_{32}(m), t_{33}(m)\}$ (m arbitrary in each generator) completely expresses the transformations of (α_{ij}) by the group P_1. Defining $T_2,\cdots,T_7$ in the same manner, we have

$$T_2=\{t_{12}(m)\cdot t_{23}(m), t_{13}(m)\};\quad T_3=\{t_{11}(m), t_{22}(m), t_{12}(m), t_{21}(m)\};$$
$$T_4=\{t_{33}(m), t_{12}(m)\};\quad T_5=T_3;\quad T_6=T_4;\quad T_7=\{t_{11}(m), t_{22}(m)\}.$$

We separate the matrices $M(x)$ having a given determinant and a given x-section $M_i(x_1, 0, x_3)$ into sets such that the members of any given set include all the matrices equivalent to any single member under the group T_i just defined, and τ. Then we shall consider the automorphisms of $f(x)$. Let $x\rightarrow x'$ be an automorphism of $f(x)$. We have then $f(x)=g(x')=mf(x')$ and hence $f(x_1, 0, x_3)=mf(x'_1, 0, x'_3)$. Let $M(x_1, x_2, x_3)\rightarrow M'(x'_1, x'_2, x'_3)$. Then $|M'(x_1, 0, x_3)|=mf(x_1, 0, x_3)$. Now $M'(x_1, 0, x_3)$ may not be one of the canonical x-sections $M_i(x_1, 0, x_3)$. If it is not we can find constant matrices P and Q such that $M''(x) = PM'(x)Q$ where $M''(x_1, 0, x_3)$ is one of the canonical x-sections. We say that $M(x)$ is replaced by $M''(x)$ under the given automorphism. The sets of solutions including $M(x)$ and $M''(x)$ are in the same τ-class, and, conversely, if two of the above sets of solutions are such that for no automorphism is any member of one replaced by any member of the other, then the two sets are in distinct τ-classes.

Summarizing, to determine the complete set of τ-classes for $l = 3$, $m = k = 3$: 1) select a representative from each projective class of ternary cubics and determine its automorphisms; 2) determine a canonical set of x-sections $M_i(x_1, 0, x_3)$ such that $| M_i(x_1, 0, x_3) | = f(x_1, 0, x_3)$ and the pairs of constant matrices P and Q such that $PM_i(x_1, 0, x_3)Q = M_i(x_1, 0, x_3)$; 3) determine the matrices (a_{ij}) such that $| M_i(x_1, 0, x_3) + (a_{ij})x_2 | = f(x)$, and separate these into sets whose members are equivalent under the pairs P, Q, and τ; 4) combine such of these sets as are equivalent under the automorphisms of $f(x)$, this final combination giving the τ-classes 3, 3, 3 (and 6, 3, 3). We shall insist throughout that no x-section be of rank zero as that implies $l < 3$.[14]

9. $l = 3$; actual determination of τ-classes (3, 3, 3). We shall now list the τ-classes (3, 3, 3) for each class or set of classes of cubics, giving the complete computation in certain typical cases, and in others merely indicating the procedure.

Class 1) $f(x) = x_1^3 - x_2^2 x_3$. The group of automorphisms, A_1, is generated by $\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1/\alpha & 0 \\ 0 & 0 & \alpha^2 \end{pmatrix}$. Here we use M_1 and M_2. With M_1 we have equations B_1 with $a_4 = a_5 = a_2 = a_{10} = 0$, $a_8 = -1$. Equation 5) gives $a_{21} = 0$ and this in 4) gives $a_{23}a_{31} = -1$ whence $a_{23}a_{31} \neq 0$. Now supposing that we had a solution of equations B_1 with $a_{32} \neq 0$, we would have after using $t_{12}(-a_{32}/a_{31})$ an equivalent solution with $a'_{32} = 0$. For

$$(a_{ij}) = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ 0 & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix}$$

---

[14] There is an obvious generalization of this paragraph for the cases $l = 3$, $m = k$ for all values of m. But for $m > 3$ the difficulties of computation become very great, especially for the first of the four steps indicated. The methods indicated here will work not only for the $GF[p]$, but for trilinear forms in any field. If, however, one is working in an algebraically closed field somewhat simpler methods will apply. The author is now working on the 3, 3, 3 case for the complex number field.

becomes

$$(a'_{ij}) = \begin{pmatrix} a_{11} & a_{12} + (a_{22} - a_{11})(a_{32}/a_{31}) & a_{13} + (a_{23}a_{32}/a_{31}) \\ 0 & a_{22} & a_{23} \\ a_{31} & 0 & a_{33} \end{pmatrix}.$$

Hence we lose no generality in supposing $a_{32} = 0$ in equations B_1. Similarly using $t_{13}(-a_{33}/a_{31})$ we have $a'_{33} = 0$. Then applying $t_{33}(1/a_{31})$ we have, in view of 4), $a'_{31} = 1 - a'_{23}$. Then $t_{32}(a_{22}) \cdot t_{13}(-a_{22})$ gives $a'_{22} = a'_{32} = a'_{33} = a'_{21} = 0$. Substituting in B_1 we have:

1) $a_{11} = 0$; 2) $-a_{13}a_{31} = 0$ whence $a_{13} = 0$; 3) $-a_{12} = 0$

giving $(a_{ij}) = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -1 \\ 1 & 0 & 0 \end{pmatrix}$ and representing the only τ-class for

$$M(x_1, 0, x_3) = M_1.$$

Now from M_2 we get equations B_2 with $a_4 = a_5 = a_2 = a_9 = a_{10} = 0$, $a_8 = -1$. Substituting 5) $a_{31} = 0$ and 6) $-a_{21} - a_{32} = 0$ in 4), we have $a_{32}(a_{33} - a_{11}) = -1$. Hence $a_{32} \neq 0$. Operating with

$$t_{12}(-a_{33}/a_{32}) \cdot t_{23}(-a_{33}/a_{32})$$

we have $a'_{33} = 0$, $a'_{11} \neq 0$. Now $t_{13}(-a_{13}/a_{11})$ gives $a'_{13} = 0$ and B_2 becomes:

1) $a_{11} + a_{22} = 0$ or $a_{22} = -a_{11}$; 2) $a_{11}a_{22} - a_{23}a_{32} + a_{12}a_{32} = 0$;
3) $-a_{11}a_{23}a_{32} = 0$, hence $a_{23} = 0$; 4) $a_{11}a_{32} = 1$;

giving

$$(a_{ij}) = m(a) = \begin{pmatrix} 1/a & 1/a^3 & 0 \\ -a & 1/a & 0 \\ 0 & a & 0 \end{pmatrix}; \quad a \neq 0.$$

Now we use the automorphisms of $f(x)$, and τ to investigate the possibility of $m(a) \overset{\tau}{\sim} m(a')$. The automorphism $\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1/\alpha & 0 \\ 0 & 0 & \alpha^2 \end{pmatrix}$, $\alpha \neq 0$, gives $m(a) \sim m(\alpha a)$ and therefore the single τ-class given by $m(1) = \begin{pmatrix} 1 & 1 & 0 \\ -1 & -1 & 0 \\ 0 & 1 & 0 \end{pmatrix}$.

Classes 2) and 3) $f(x) = x_1^3 + x_1^2x_2 - \gamma x_3^2x_2$. For class 2) $\gamma = 1$,

$$A_2 = \left\{ \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{pmatrix}, \begin{pmatrix} c & 0 & 1 \\ -9c/4 & -2c & -3/4 \\ -1 & 0 & c \end{pmatrix} \text{ where } c^2 = -1/3 \right\};$$

and for class 3), $\gamma \neq$ square,

$$A_3 = \left\{ \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{pmatrix}, \begin{pmatrix} 4 & 0 & 12 \\ -9 & -8 & -9 \\ 4 & 0 & -4 \end{pmatrix} \text{ where } p = 6k + 5, \gamma = -3 \right\}.$$

$f(x_1, 0, x_3) = x_1^3$ and since there are terms in x_3^2 in $f(x)$ we must have $M(0, 0, 1)$ of rank two; hence we need only consider M_2 giving B_2 with $a_4 = 1$, $a_9 = -\gamma$, $a_5 = a_2 = a_8 = a_{10} = 0$. There are for each of these two classes of cubics $(p + 3)/2$ τ-classes represented by $m_1(a) = \begin{pmatrix} 0 & 0 & 0 \\ -a & 1 & 0 \\ -\gamma & a & 0 \end{pmatrix}$, $m(a) \overset{\tau}{\sim} m(-a)$; and $m_2 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ -\gamma & 0 & 0 \end{pmatrix}$.

Class 4) $f(x) = x_1^3 + \alpha x_2^3 + x_1 x_2 x_3$, $\alpha \neq$ cube, whence $p = 6k + 1$. Campbell[15] lists this as set 4) giving a cubic with parameter α as equivalent to another with parameter $\alpha' = k^3\alpha$ for any k. He failed to consider interchanging the two tangents at the double point, which gives cubics with parameters α and α' in the same class if $\alpha' = k^3\alpha^2$. But if $\alpha \neq$ cube, any other not-cube in the $GF[p]$ is one of the forms $k^3\alpha$ or $k^3\alpha^2$ for some k. A_4 is generated by $\begin{pmatrix} 1 & 0 & 0 \\ 0 & \omega & 0 \\ 0 & 0 & \omega^2 \end{pmatrix}$, $\omega^3 = 1$. x_3^2 does not appear in $f(x)$; so we have cases with both M_1 and M_2. From M_1 we get the single τ-class represented by $m_1 = \begin{pmatrix} 0 & 0 & -\alpha \\ -1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}$, and from M_2 the $(p - 1)/2$ τ-classes represented by $m_2(a) = \begin{pmatrix} 0 & 0 & -\alpha/a(1 + a) \\ -1 - a & 0 & 0 \\ 0 & a & 0 \end{pmatrix}$, $a \neq 0, -1$; $m_2(a) \overset{\tau}{\sim} m_2(-a)$.

Class 5) $f(x) = x_1^3 + \alpha x_2^3 + x_1^2 x_3 + 3x_2^2 x_3$, $p = 6k + 5$. A is generated by $\begin{pmatrix} 1 & 3 & 0 \\ -1 & 1 & 0 \\ (\alpha - 9)/4 & (-3\alpha - 9)/4 & -2 \end{pmatrix}$. Campbell's result differs somewhat from this due to a mistake in calculation. On p. 35[16] the flex equation should be $\gamma\alpha c^3 - 3c^2 + 3\gamma^2\alpha c - \gamma = 0$ instead of $\gamma\alpha c^3 + 3c^2 - 3\gamma^2\alpha c - \gamma = 0$ as it is there. Then his conditions on α and p that the set exist become $-(3/\gamma) =$ square or since $\gamma \neq$ square; $-3 \neq$ square, $p = 6k + 5$; and

---

[15] *Loc. cit.*, p. 35.

[16] *Loc. cit.*

$\alpha \neq \beta\left(\frac{\beta^2-1}{9\beta^2-1}\right)$ for any β in the $GF[p]$. (Since γ is any particular not-square we take $\gamma = -3$.) Further $1 + \alpha\gamma^{3/2}$, $\alpha \neq 0$, is a cube in the $GF[p^2]$ for only $(p-5)/3$ values of α. The other $(2p+2)/3$ will not give cubes and are therefore not of the form $\beta\left(\frac{\beta^2-1}{9\beta^2-1}\right)$.

The reducibility of $g(x, \alpha'') = \gamma x^3 + 3\gamma^2\alpha'' x + 3x + \gamma\alpha''$, $\alpha'' = \frac{\alpha'-\alpha}{\alpha\alpha'\gamma^3-1}$, is the condition that $f(x, \alpha)$ be equivalent to $f(x, \pm\alpha')$ i. e. if using either one of α' and $-\alpha'$ gives the cubic $g(x, \alpha'')$ reducible, then $f(x, \alpha)$ is equivalent to both $f(x, \alpha')$ and $f(x, -\alpha')$. But the reducibility of $g(x, \alpha'')$ implies $\alpha'' = \beta\left(\frac{\beta^2-1}{9\beta^2-1}\right)$ for some β. That is, α'' is one of the set B of $(p+1)/3$ numbers (including 0 and ∞) giving cubics $f(x, \alpha'')$ with flexes; conversely, if α'' is any such number, from $\alpha'' = \frac{\alpha'-\alpha}{\alpha\alpha'\gamma^3-1}$ we have $\alpha' = \frac{\alpha''-\alpha}{\alpha\alpha''\gamma^3-1}$ where $f(x, \alpha')$ is equivalent to $f(x, \alpha)$. As α'' takes its $(p+1)/3$ possible values, α' takes $(p+1)/3$ distinct values (including α). Now if $-\alpha'$ is not in this set for each (or any) α' we have $(2p+2)/3$ distinct numbers giving equivalent cubics $f(x, \pm\alpha')$. But the set A of numbers α giving cubics $f(x, \alpha)$ without flexes contains just $(2p+2)/3$ numbers. Hence, in this case all of the cubics with acnodes but no (real) flex points belong to one projective class. We shall show that $-\alpha'$ and α' are never in the same set defined by $\frac{\alpha''-\alpha}{\alpha\alpha''\gamma^3-1}$ for fixed α in A, α'' varying, as above, in B and hence that there is always just one class.

For suppose that $\alpha' = \frac{\alpha''-\alpha}{\alpha\alpha''\gamma^3-1}$ and $-\alpha = \frac{\beta''-\alpha}{\alpha\beta''\gamma^3-1}$, α'' and β'' in B. Then

$$\beta'' = \frac{\alpha'' - \left(\frac{2\alpha}{1+\alpha^2\gamma^3}\right)}{\left(\frac{2\alpha}{1+\alpha^2\gamma^3}\right) - 1}$$

which implies that $2\alpha/(1+\alpha^2\gamma^3)$ is in B. Further since α can be taken as any one of the $(2p+2)/3$ numbers in the set A, $2\alpha/(1+\alpha^2\gamma^3)$ must be in B for every α in A. But $a = 2\alpha/(1+\alpha^2\gamma^3)$ cannot have more than two solutions α for any a. Hence as α takes every value in A, $2\alpha/(1+\alpha^2\gamma^3)$ must take at least $(p+1)/3$ distinct values, and being always in B must fill it. But since 0 is in B this implies $0 = 2\alpha/(1+\alpha^2\gamma^3)$, or $\alpha = 0$ or $\alpha = \infty$ for α in A. But 0, ∞ are not in A, giving a contradiction to the supposition that

α' and $-\alpha'$ had the above representations, which completes the proof that all such cubics are projectively equivalent. $f(x_1, 0, x_3) = x_1^2(x_1 + x_3)$ so we need to consider both M_3 and M_4. From M_3 we have the single solution $m_1 = \begin{pmatrix} 0 & 1 & 3 \\ -3 & 0 & \alpha \\ 1 & 0 & 0 \end{pmatrix}$, and from M_4 the $(p+1)/2$ τ-classes given by

$$m_2(a) = \begin{bmatrix} -a & \dfrac{\alpha - a^3}{3 + a^2} & -a^2 \\ 0 & a & 3 + a^2 \\ 1 & 0 & 0 \end{bmatrix}, \; m_2(a) \overset{\tau}{\sim} m_2(-a).$$

Sets 6) and 7) $f(x) = x_1(x_1 - x_2)(x_1 - \alpha x_2) - \gamma x_2 x_3^2$; $\gamma = 1$ (set 6), $\gamma \neq$ square (set 7). $A_6 = A_7 = \left\{ \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{pmatrix}, \begin{pmatrix} 1 & -1 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & \sqrt{-1} \end{pmatrix} \text{ if } \alpha = 1/2, \begin{pmatrix} 1 & \omega & 0 \\ 0 & \omega & 0 \\ 0 & 0 & \omega \end{pmatrix} (\omega^3 = 1) \text{ if } \alpha^2 - \alpha - 1 = 0 \right\}$. Here we need only consider M_2 giving the solutions $m(a, b) = \begin{pmatrix} -1 - \alpha - a & b & -b^2/a \\ 0 & a & -b \\ -\gamma & 0 & 0 \end{pmatrix}$ where $b = (-a/\gamma)(\alpha + a)(1 + a)$ and of course values of a for which b would not be real are discarded. For all values of α, $m(a, b) \overset{\tau}{\sim} m(a, -b)$; if $\alpha = 1/2$ and $p = 4k + 1$, $m(a, b) \overset{\tau}{\sim} m(-1 - a, b)$. If $\alpha^2 - \alpha + 1 = 0$, $m(a, b) \overset{\tau}{\sim} m(\omega a + \omega, b)$. Aside from these cases any two values of a for which b is real give distinct τ-classes.

Sets 8) and 9) $f(x) = x_1(x_1^2 - \alpha x_1 x_2 + \alpha x_2^2) - \gamma x_2 x_3^2$, $\alpha^2 - 4\alpha \neq$ square, $\gamma = 1$ (set 8), $\gamma \neq$ square (set 9). $A_8 = A_9 = \left\{ \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{pmatrix} \right\}$. We have only M_2 to consider giving the solutions $m(a, b) = \begin{pmatrix} -\alpha - a & b & -b^2/a \\ 0 & a & -b \\ -\gamma & 0 & 0 \end{pmatrix}$, $\gamma b^2 = -a(a^2 + \alpha a + \alpha)$, $m(a, b) \overset{\tau}{\sim} m(a, -b)$. There is a τ-class for each a giving b real.

Set 8a) (two classes) $f(x) = x_1(x_1^2 - \alpha x_2^2) - x_2 x_3^2$, $\alpha = \gamma$ or $1/\gamma$, $\gamma \neq$ square. (Campbell omits this set.) $A_{8a} = \left\{ \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{pmatrix}, \begin{pmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & \sqrt{-1} \end{pmatrix} \right\}$. The τ-classes are given by $m(a, b) = \begin{pmatrix} -a & b & a^2 - \alpha \\ 0 & a & b \\ -1 & 0 & 0 \end{pmatrix}$, $b^2 = -a(a^2 - \alpha)$, with $m(a,b) \overset{T}{\sim} m(a,-b)$ and if $-1 =$ square $m(a,b) \overset{T}{\sim} m(-a, \sqrt{-1}\,b)$ Otherwise values of a for which b is real give distinct τ-classes.

Sets 10) and 11) $f(x) = x_1^3 - \alpha x_1 x_2^2 + \alpha x_2^3 - \gamma x_2 x_3^2$, $\gamma = 1$ (set 10), $\gamma \neq$ square (set 11). $A_{10} = A_{11} = A_8$. The τ-classes are given by

$$m(a,b) = \begin{pmatrix} -a & b & (a^2-\alpha)/\gamma \\ 0 & a & -b \\ -\gamma & 0 & 0 \end{pmatrix},\ \gamma b^2 = a^3 - \alpha a + \alpha,\ m(a,b) \overset{T}{\sim} m(a,-b).$$

Set 12) $f(x) = x_1^3 - \alpha x_2^3 - x_2 x_3^2$, $\alpha \neq$ cube.

$$A_{12} = \left\{ \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{pmatrix}, \begin{pmatrix} \omega & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} (\omega^3 = 1) \right\}.$$

The τ-classes are given by $m(a,b) = \begin{pmatrix} -a & b & a^2 \\ 0 & a & -b \\ -1 & 0 & 0 \end{pmatrix}$, $b^2 = \alpha - a^3$, where $m(a,b) \overset{T}{\sim} m(\omega a, -b)$.

Sets 13) and 16) $f(x) = x_1^3 - x_2 x_3^2 + \alpha x_2^2 x_3$.

$$A_{13} = \left\{ \begin{pmatrix} \omega & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} (\omega^3 = 1), \begin{pmatrix} \alpha & 0 & 0 \\ 0 & 0 & -1 \\ 0 & -\alpha^2 & 0 \end{pmatrix}, \begin{pmatrix} -\alpha^2 & 0 & 1 \\ 0 & 0 & 1 \\ 0 & -\alpha^2 & \alpha \end{pmatrix} \right\}.$$

A_{16} same as A_{13} except that there may be additional automorphisms obtained by taking for $x'_1 = 0$ any other line of flexes. (Such transformation will give either automorphisms or equivalent cubics in the same form but with different values of α). The τ-classes are given by $m(a,b) = \begin{pmatrix} -a & b & a^2 \\ 0 & a & -b-\alpha \\ 1 & 0 & 0 \end{pmatrix}$, $b^2 + \alpha b + a^3 = 0$, where the two roots b for a given a give equivalent solutions, and $m(a,b) \sim m(\omega a, b)$.

Sets 14) and 15) $f(x) = x_1^3 + x_2 x_3 (x_1 - \alpha x_2 - x_3)$.

$$A_{14} = \left\{ \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1/\alpha \\ 0 & \alpha & 0 \end{pmatrix}, \begin{pmatrix} 1 & 0 & 0 \\ 1/\alpha & -1 & -(1/\alpha) \\ 0 & 0 & 1 \end{pmatrix} \right\}.$$

$A_{15} = A_{14}$ plus automorphisms that might arise from taking for $x'_1 = 0$ another of the twelve flex lines. Campbell's discussion as to values of α giving set 14), i. e. just three real flex points, is incomplete. He gives[17] sufficient but not necessary conditions on α, and makes a numerical error in the calculations that he does give. He solves $x = ay + bz$ simultaneously with

---

[17] *Loc. cit.*, p. 43.

$f(x, y, z) = \alpha z^3 + xy(z - x - y) = 0$ and requires that the result be a perfect cube. This gives a rationally in terms of b and c where $b^2 - b + 3\alpha c = 0$ and c is a root of $3\alpha c^3 - c^2 - 3c - 3 = 0$. The discriminant of this cubic is a square multiple of -3. Now if a and b are real the curve $f(x, y, z) = 0$ has four, and therefore nine, real flex points. But then there must be six distinct real solutions b, which implies that the cubic in c has three real zeros. But the cubic has three real zeros only when $-3 =$ square and α is of the form $(c^2 + 3c + 3)/3c^3$.[18] Otherwise, i. e. $-3 \neq$ square or $-3 =$ square, $\alpha \neq (c^2 + 3c + 3)/3c^3$, the curve $f(x, y, z)$ has just three real flex points. The representation given above is readily derived from this one. The τ-classes are given by $m(a, b) = \begin{pmatrix} -a & b & a^2 - b \\ -1 & a & \alpha - b \\ -1 & 0 & 0 \end{pmatrix}$, $b^2 - (a + \alpha)b + a^3 = 0$, the two roots b for a given a giving equivalent τ-classes.

Sets 17), 18), 19) $f(x) = x_1{}^2x_3 + x_1x_2{}^2 + \alpha x_2x_3{}^2 + ax_1{}^2x_2 + bx_1x_2x_3$, $a = 1$, $b = -\beta$ (set 17); $a = 0$, $b = 1$ (set 18); $a = b = 0$ (set 19). For sets 17), 18), and 19) the computations involved in determining the classes and then the complete group of automorphisms seem too involved to be worth doing. For sets 18) and 19) we list the automorphism $\begin{pmatrix} 0 & \alpha & 0 \\ 0 & 0 & \alpha \\ 1 & 0 & 0 \end{pmatrix}$ and for set 19) the further automorphism $\begin{pmatrix} 1 & 0 & 0 \\ 0 & \omega & 0 \\ 0 & 0 & \omega^2 \end{pmatrix}$. The τ-classes for set 17) are given by $m_1(a, b) = \begin{pmatrix} -\beta & -b/\alpha - 1 & -ab - 1 - \beta \\ -\alpha & 0 & b \\ 1 & a & 1 \end{pmatrix}$, $(\alpha b)a^2 + (\alpha + \alpha\beta + \beta b)a - (1/\alpha)(\alpha + b)^2 = 0$, the two roots a for given b giving equivalent solutions. For set 18) $m_2(a, b) = \begin{pmatrix} 1 & -b/\alpha & -ab - 1 \\ -\alpha & 0 & b \\ 1 & a & 0 \end{pmatrix}$, $(\alpha b)a^2 + (\alpha - b)a - b^2/\alpha = 0$, the two roots a for given b giving equivalent solutions. For set 19) $m_3(a, b) = \begin{pmatrix} 0 & -b/\alpha & -ab - 1 \\ -\alpha & 0 & b \\ 1 & a & 0 \end{pmatrix}$,

$$(\alpha b)a^2 + \alpha a - b^2/\alpha = 0,$$

the two roots a for given b giving equivalent solutions and also

$$m_3(a, b) \overset{\tau}{\sim} m_3(\omega a, \omega^2 b).$$

---

[18] L. E. Dickson, *Bulletin of the American Mathematical Society*, vol. 13 (1906), pp. 1-8.

Class 20) A cubic without real points. L. E. Dickson[19] proved the existence of such cubics showing that they were all conjugate under the collineation group and that each such cubic consists of three imaginary lines. We give here another proof of the existence of cubics of this class.

Consider any binary cubic $F(x) = x_1^3 - a_1 x_1^2 x_2 - a_2 x_1 x_2^2 - a_3 x_2^3$ belonging to and irreducible in the $GF[p]$. The matrix $T = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ a_3 & a_2 & a_1 \end{pmatrix}$ has for its characteristic equation $-F(x, 1) = 0$. *The matrices in the* $GF[p]$ *permutable with* T *are all powers of some matrix whose characteristic determinant is a primitive irreducible cubic, and of which* T *is a power.*[20] This can be seen by transforming T into the irrational canonical form $T^* = \begin{pmatrix} \rho & 0 & 0 \\ 0 & \rho^p & 0 \\ 0 & 0 & \rho^{p^2} \end{pmatrix}$ where ρ is a root of $F(x, 1) = 0$. If σ is a primitive mark in the $GF[p^3]$ we have $\rho = \sigma^n$. Then $T^* = \begin{pmatrix} \sigma & 0 & 0 \\ 0 & \sigma^p & 0 \\ 0 & 0 & \sigma^{p^2} \end{pmatrix}^n$ and is therefore permutable with the powers of $T^*_\sigma = \begin{pmatrix} \sigma & 0 & 0 \\ 0 & \sigma^p & 0 \\ 0 & 0 & \sigma^{p^2} \end{pmatrix}$. There exists a matrix S such that $T_\sigma = S^{-1} T^*_\sigma S$ is in the $GF[p]$. Then $S^{-1} T^* S$ is likewise in the $GF[p]$ and has for its rational canonical form, T, and so there exists S' such that $S'^{-1}(S^{-1} T^* S) S' = T$. Then the $p^3 - 1$ distinct powers of $S'^{-1} T_\sigma S'$ are permutable with T. We complete the proof by showing that there are just $p^3 - 1$ matrices permutable with T and will further exhibit their form. If $(\alpha_{ij}) T = T(\alpha_{ij})$ we have

$$\begin{pmatrix} a_3\alpha_{13} & \alpha_{11} + a_2\alpha_{13} & \alpha_{12} + a_1\alpha_{13} \\ a_3\alpha_{23} & \alpha_{21} + a_2\alpha_{23} & \alpha_{22} + a_1\alpha_{23} \\ a_3\alpha_{33} & \alpha_{31} + a_2\alpha_{33} & \alpha_{32} + a_1\alpha_{33} \end{pmatrix}$$

$$= \begin{pmatrix} \alpha_{21} & \alpha_{22} & \alpha_{23} \\ \alpha_{31} & \alpha_{32} & \alpha_{33} \\ a_3\alpha_{11} + a_2\alpha_{21} + a_1\alpha_{31} & a_3\alpha_{12} + a_2\alpha_{22} + a_1\alpha_{32} & a_3\alpha_{13} + a_2\alpha_{23} + a_1\alpha_{33} \end{pmatrix}$$

giving $(\alpha_{ij}) = \alpha_{11} I + \alpha_{12} T + \alpha_{13} T^2$, where the α_{1j}'s are arbitrary (in the $GF[p]$). There are just $p^3 - 1$ such matrices and they are therefore the $p^3 - 1$ powers of the matrix $S'^{-1} T_\sigma S'$.

Now consider

---

[19] L. E. Dickson, *Bulletin of the American Mathematical Society*, vol. 14, ser. 2 (1908), pp. 160-169.

[20] Compare with Jordan, *Traité de Substitutions* (1870), 128 ff.

$$\begin{aligned} f(x) = | x_1I + x_2T + x_3T^2 | &= x_1^3 + a_1x_1^2x_2 - a_2x_1x_2^2 + a_3x_2^3 \\ &+ a_1a_3x_2^2x_3 - a_2a_3x_2x_3^2 + a_3^2x_3^3 + (2a_2 + a_1^2)x_1^2x_3 \\ &+ (a_2^2 - a_1a_3)x_1x_3^2 + (-3a_3 - a_1a_2)x_1x_2x_3. \end{aligned}$$

$f(x_1, x_2, x_3) = 0$ implies $x_1 = x_2 = x_3 = 0$, for otherwise some power of a non-singular matrix would have a zero determinant.

Now suppose that $F(x)$ is primitive. Then $M(x) = x_1I + x_2T + x_3T^2$ gives a form whose determinant is the imaginary cubic. We now search for the general solution of $| x_1I + x_2T + x_3(a_{ij}) | = f(x)$. We get the following equations, B_8:

1) $a_{11} + a_{22} + a_{33} = 2a_2 + a_1^2$ 3) $| a_{ij} | = a_3^2$
2) $a_{11}a_{22} + a_{11}a_{33} + a_{22}a_{33} - a_{12}a_{21} - a_{13}a_{31} - a_{23}a_{32} = a_2^2 - a_1a_3$
4) $-a_{21}a_{33} + a_{23}a_{31} + a_{12}a_{31} - a_{11}a_{32} + a_3a_{12}a_{23} - a_3a_{13}a_{22} + a_2a_{13}a_{21}$
$- a_2a_{11}a_{23} + a_1a_{11}a_{22} - a_1a_{12}a_{21} = - a_2a_3$
5) $a_{11} + a_3a_{12} + a_3a_{23} - a_2a_{11} - a_1a_{21} = a_1a_3$
6) $a_1a_{22} - a_2a_{23} - a_{32} - a_{21} - a_3a_{13} = -3a_3 - a_1a_2$.

Now eliminating a_{33}, a_{32}, a_{31} from 2), 3), and 4) by means of 1), 5), and 6) we have three conditions remaining, two quadratic and one cubic, on the six elements of the first two rows of (a_{ij}). We cannot have $a_{11} = a_{12} = a_{13} = 0$. For any other of the p^3 sets of values for a_{11}, a_{12} and a_{13} we have at most twelve solutions of 2), 3), and 4) in a_{21}, a_{22}, and a_{23} and hence at most $12(p^3 - 1)$ solutions for the system of equations.

Now let $A = (a_{ij})$ represent any solution of B_8 not a power of T. Then $T^{-\alpha}(x_1I + x_2T + x_3A)T^{\alpha}$ gives an equivalent solution $(x_1I + x_2T + x_3T^{-\alpha}AT^{\alpha})$. Since A is not a power of T and since A belongs to the exponent $(p^3 - 1)/2$ (having the same characteristic polynomial as T^2), the equation $T^{-\alpha}AT^{\alpha} = A$ implies $\alpha = k(p^2 + p + 1)$. Hence the solutions A, not powers of T, can be grouped into sets of $p^2 + p + 1$, the members of any set being in the same τ-class.

Now if A is a power of T, say $A = T^{\alpha}$, since it has the same characteristic polynomial as T^2 we must have $A = T^2$, T^{2p} or T^{2p^2}, for if a matrix P has an irreducible characteristic polynomial the only ones of its powers conjugate to it (i. e. having the same characteristic polynomial) are P^{p^β}, $\beta = 0, 1, \cdots$.

Now it is readily verified that $T' = \begin{pmatrix} 0 & 0 & a_3 \\ 1 & 0 & a_2 \\ 0 & 1 & a_1 \end{pmatrix}$ is an automorphism of $f(x)$ replacing $f(x)$ by $a_3f(x')$, and our solution by

$$(x'_1T + x'_2A + x'_3(a_3I + a_2T + a_1A)).$$

Now multiply on the right by T^{-1} giving

$$(x'_1I + x'_2AT^{-1} + x'_3[a_3T^{-1} + a_2I + a_1AT^{-1}]).$$

Now

$$|\, x_1I + x_2T \,| = |\, x_1I + x_2AT^{-1} \,|$$

so if A is a power of T, AT^{-1} must be conjugate to T and hence $AT^{-1} = T$, T^p or T^{p^2} giving $A = T^2$, T^{p+1} or T^{p^2+1}. This combined with the above restrictions on A gives $A = T^2$.

Hence, we have at most $\dfrac{12(p^3-1)}{p^2+p+1} = 12(p-1)$ τ-classes with the determinant $f(x)$. Of these we have proved the existence of at least one.

It is worth mentioning here that the scheme of classification being followed gives complete results in all cases save sets 15-19 (where we listed all possible solutions but could not prove them all distinct) and class 20 (where we have been unable to list all possible solutions).

Class 21) $f(x) = x_1^3 + x_1x_2x_3$ gives M_1 and M_2. (We shall not list the automorphisms for the reducible cubics.) The τ-classes are:

$$m_1 = \begin{pmatrix} 0 & 0 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}; \qquad m_2 = \begin{pmatrix} 0 & 0 & 1 \\ -1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

from $M_1(x_1, 0, x_3)$ and

$$m_3(a) = \begin{pmatrix} 0 & 0 & 0 \\ a & 0 & 0 \\ 0 & -1-a & 0 \end{pmatrix}, \; a(a+1) \neq 0, \; m_3(a) \overset{\tau}{\sim} m_3(-1-a);$$

$$m_4(a) = \begin{pmatrix} 0 & 0 & a \\ -1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \; a \neq 0, \; m_4(a) \sim m_4(\alpha^3 a) \sim m(\alpha^3 a^2),$$

α arbitrary, from $M_2(x_1, 0, x_3)$.

Class 22) $f(x) = x_1(x_1^2 - x_3^2 + \gamma x_2^2)$. The τ-classes (all from $M_7(x_1, 0, x_3)$) are:

$$m_1(a) = \begin{pmatrix} 0 & 1 & a \\ 0 & 0 & -\gamma \\ 0 & 1 & 0 \end{pmatrix}, \; m_1(a) \overset{\tau}{\sim} m_1(-a), \; m_1(a) \overset{\tau}{\sim} m_1(a')$$

where

$$a' = \frac{b(2c+1)a - (2c-1)(c+1)}{(c+1)(2c-1)(a/\gamma) - b(2c+1)}, \; \gamma b^2 = c^2 - 1;$$

$$m_2 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -\gamma \\ 0 & 1 & 0 \end{pmatrix}; \quad m_3(a) = \begin{bmatrix} 0 & \dfrac{-\gamma + a^2}{2} & \dfrac{-\gamma + a^2}{2} \\ 1 & 0 & a \\ 1 & -a & 0 \end{bmatrix}, \; m_3(a) \overset{\tau}{\sim} m_3(-a).$$

The values a such that $m_1(a) \sim m_1(0)$ are given by $a = \dfrac{(2c-1)(c+1)}{b(2c+1)}$. Eliminating b from this and $\gamma b^2 = c^2 - 1$ we get the cubic equation $C)$ $c^3 - \frac{3}{4}c - \frac{1}{4}\left(\dfrac{a^2+\gamma}{a^2-\gamma}\right) = 0$. For $m_1(a) \sim m_1(0)$ it is necessary and sufficient that $C)$ have a root c such that b is real. The discriminant of $C)$ is $(-3\gamma)\left[\dfrac{2(a^2-\gamma)}{3a}\right]^2$. If it is a not-square i. e. if $-3 =$ square $(p = 6k+1)$ $C)$ will have one and just one real root, c, for every $a \neq 0$, a and $-a$ giving the same root.[21] Then the values of a for which b is real and different from zero[22] will give $p-1$ values $a \neq 0$ for which $m_1(a) \sim m_1(0)$ and therefore just one τ-class for $p = 6k+1$.

If $m_1(a) \sim m_1(a')$ for more than one value of c we have

$$\frac{b(2c+1)a - (2c-1)(c+1)}{(c+1)(2c-1)(a/\gamma) - b(2c+1)} = \frac{b'(2c'+1)a - (2c'-1)(c'+1)}{(c'+1)(2c'-1)(a/\gamma) - b'(2c'+1)}$$

which implies

$$\frac{(2c-1)(c+1)}{b(2c+1)} = \frac{(2c'-1)(c'+1)}{b'(2c'+1)}$$

or in other words that $C)$ shall have two and therefore three real roots. Furthermore it is evident that if c, b, and c' are real, then b' must also be real, and hence that if $C)$ has three real roots either none or all of them give b real. Thus for $-3 \neq$ square the $(p-1)/2 - 2$ values of c aside from $\pm\frac{1}{2}$ for which b is real and different from zero will give $(p-5)/3$ distinct values $a \neq 0$ for which $m_1(a) \sim m_1(0)$. For $a \neq 0$: $c = 1$; $c = -\frac{1}{2}$, $b = \pm\sqrt{-3/4\gamma}$ give $m_1(a) \sim m_1(-\gamma/a)$, and $c = -1$; $c = \frac{1}{2}$, $b = \pm\sqrt{-3/4\gamma}$ give $m_1(a) \sim m_1(-a)$. Now suppose that $\bar{a}$ is not one of the $(p-2)/3$ values a (including 0) for which $m_1(a) \sim m_1(0)$. Then the above mentioned $(p-1)/2 - 2$ values of c for which b is real give $(p-5)/3$ distinct values $\bar{a}'$ which with $-\bar{a}$ and $\gamma/-\bar{a}$ give

$$(p-5)/3 + 2 = (p+1)/3$$

values a for which $m_1(\bar{a}') \sim m_1(\bar{a})$. Since

$$(p-2)/3 + (p+1)/3 + (p+1)/3 = p$$

there are for $p = 6k+5$ just three τ-classes given by the solutions $m_1(a)$.

---

[21] L. E. Dickson, *Bulletin of the American Mathematical Society*, vol. 13 (1906), 1 ff.

[22] L. E. Dickson, *Linear Groups*, Chapter IV.

Class 23) $f(x) = x_1{}^2x_3 + x_3{}^2x_2$. The τ-classes (all from $M_6(x_1, 0, x_3)$ are given by

$$m_1 = \begin{pmatrix} 0 & 0 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}; \qquad m_2 = \begin{pmatrix} 0 & 0 & 0 \\ -1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix};$$

$$m_3 = \begin{pmatrix} 0 & 0 & 0 \\ -1 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix}; \qquad m_4 = \begin{pmatrix} 0 & -1 & -1 \\ -1 & 0 & 1 \\ 1 & 1 & 0 \end{pmatrix}.$$

Class 24) $f(x) = x_1x_2x_3$. The τ-classes are given by:

$$(1) \begin{pmatrix} x_1 & 0 & 0 \\ 0 & x_2 & 0 \\ 0 & 0 & x_3 \end{pmatrix}; \qquad (2) \begin{pmatrix} x_1 & 0 & 0 \\ 0 & x_2 & x_1 \\ 0 & 0 & x_3 \end{pmatrix}; \qquad (3) \begin{pmatrix} x_1 & 0 & x_2 \\ 0 & x_2 & x_1 \\ 0 & 0 & x_3 \end{pmatrix};$$

$$(4) \begin{pmatrix} x_1 & 0 & 0 \\ 0 & x_2 & x_1 \\ x_2 & 0 & x_3 \end{pmatrix}; \qquad (5) \begin{pmatrix} x_1 & x_3 & 0 \\ 0 & ax_2 & x_1 \\ (1-a)x_2 & 0 & x_3 \end{pmatrix}, \quad a \neq 0, 1; \ a \text{ and } (1-a)$$

giving the same τ-class;

$$(6) \begin{pmatrix} x_1 & x_3 & x_2 \\ 0 & x_2 & x_1 \\ 0 & 0 & x_3 \end{pmatrix}.^{23}$$

Class 25) $f(x) = x_1{}^2x_3 - \gamma x_2{}^2x_3$, $\gamma \neq$ square. The τ-classes are given by:

$$m_1 = \begin{pmatrix} 0 & \gamma & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}; \text{ and } \ m_2 = \begin{pmatrix} 0 & \gamma & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix} \text{ from } M_5(x_1, 0, x_3);$$

and

$$m_3(a) = \begin{pmatrix} -a & 0 & 0 \\ 0 & a & a^2-\gamma \\ 1 & 0 & 0 \end{pmatrix}, \ m_3(a) \overset{\tau}{\sim} m_3(-a), \text{ from } M_6(x_1, 0, x_3).$$

Class 26) $f(x) = x_1(x_1 + x_3)(x_1 - x_3)$. The τ-classes (all from $M_7(x_1, 0, x_3)$) are given by

$$m_1 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix}; \qquad m_2 = \begin{pmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix};$$

$$m_3 = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix}; \qquad m_4 = \begin{pmatrix} 0 & \frac{1}{2} & \frac{1}{2} \\ 1 & 0 & 1 \\ 1 & -1 & 0 \end{pmatrix}.$$

---

[23] For the computation in this class we used $f_1(x) = x_1x_3(x_1 + x_2)$ and then transformed the results back to $f(x) = x_1x_2x_3$.

Class 27) $f(x) = x_1(x_1^2 - \gamma x_2^2)$, $\gamma \neq$ square. We have two τ-classes:

$$m_1 = \begin{pmatrix} 0 & 0 & \gamma \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix} \text{ from } M_1(x_1, 0, x_3), \text{ and } m_2 = \begin{pmatrix} 0 & 0 & 0 \\ -1 & 0 & \gamma \\ 0 & 1 & 0 \end{pmatrix} \text{ from } M_2(x_1, 0, x_3).$$

Class 28) $f(x) = x_1^2(x_1 + x_2)$. The τ-classes are given by:

$$m_1 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}; \quad m_2 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}; \quad m_3 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix};$$

$$m_4 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix}; \quad m_5 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix};$$

from $M_1(x_1, 0, x_3)$; and

$$m_6 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}; \quad m_7 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix}; \quad m_8 = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix};$$

$$m_9 = \begin{pmatrix} 0 & 0 & \gamma \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix}; \quad m_{10} = \begin{pmatrix} 0 & 0 & 0 \\ -1 & 1 & 0 \\ 0 & 1 & 0 \end{pmatrix}; \text{ from } M_2(x_1, 0, x_3).$$

Class 29) $f(x) = x_1^3$. The τ-classes are given by:

$$m_1 = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}; \quad m_2 = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}; \quad m_3 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix};$$

from $M_1(x_1, 0, x_3)$; and

$$m_4 = \begin{pmatrix} 0 & 0 & 0 \\ -1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}$$

from $M_2(x_1, 0, x_3)$.

Class 30) $f(x) = x_1^3 - \alpha x_1 x_2^2 + \alpha x_2^3$ an irreducible cubic.[24] There is no τ-class from $M_1(x_1, 0, x_3)$; from $M_2(x_1, 0, x_3)$ we have the single τ-class given by $m_1 = \begin{pmatrix} 0 & 0 & -\alpha \\ -1 & 0 & -\alpha \\ 0 & 1 & 0 \end{pmatrix}$.

---

[24] Campbell's class 31) is included in class 30).

In treating $f(x) \equiv 0$ we may obtain forms $F(x, y, z)$ which have $l = 3$ but one or both of m and k less than 3. We may in such cases suppose $m \geqq k$. No such case will give a group G with $l = 3$, $m = 3$, $k = 3$, but it might give a group G with $l = 6$, $m = 3$, $k = 3$. The groups G for $k < 3$ have been classified.[25] Interpreting these results for the forms here under consideration we get: for 3, 3, 1 one τ-class represented by $\begin{pmatrix} x_1 & x_2 & x_3 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$ which does not give a group G for either $l = 3$ or $l = 6$; for 3, 2, 2 two τ-classes represented by $\begin{pmatrix} x_1 & x_3 & 0 \\ 0 & x_2 & 0 \\ 0 & 0 & 0 \end{pmatrix}$ and $\begin{pmatrix} x_1 & x_2 & 0 \\ x_3 & x_2 & 0 \\ 0 & 0 & 0 \end{pmatrix}$; for 3, 3, 2 we get eight forms which can be derived from those for $M_1, \cdots, M_7$ and M_{10} (§ 7) by interchanging the variables x and z.

For $l = m = k = 3$ we list the following:

$$M^{(0)}(x) = \begin{pmatrix} 0 & x_1 & x_2 \\ x_1 & 0 & x_3 \\ x_2 & -x_3 & 0 \end{pmatrix}, \quad M^{(1)}(x) = \begin{pmatrix} x_1 & x_2 & x_3 \\ x_2 & 0 & 0 \\ x_3 & 0 & 0 \end{pmatrix}, \quad M^{(2)}(x) = \begin{pmatrix} 0 & x_1 & x_2 \\ x_1 & 0 & 0 \\ x_3 & 0 & 0 \end{pmatrix}$$

$M^{(\nu)}(x)$ has ν x-sections of rank one, $\nu = 0, 1, 2$, and the three τ-classes are therefore distinct. We now show that these are the only such τ-classes. If $M(x)$ has no x-section of rank one and $|M(x)| \equiv 0$, $M(x_1, x_2, 0)$ may evidently be taken as M_8, M_9 or M_{10} for $l = 2$. If

$$[M_8] : \left| \begin{pmatrix} x_1 & x_2 & 0 \\ \gamma x_2 & x_1 & 0 \\ 0 & 0 & 0 \end{pmatrix} + (a_{ij})x_3 \right| = 0$$

we may by the automorphism $x'_1 = a_{11}x_3 + x_1$; $x'_2 = a_{12}x_3 + x_2$ obtain $a'_{11} = a'_{12} = 0$. Since $x_1{}^2x_3$ cannot appear $a_{33} = 0$. For $x_1x_3{}^2$ and $x_2x_3{}^2$ not to appear we must have $a_{23}a_{32} + a_{13}a_{31} = 0$ and $\gamma a_{13}a_{32} + a_{23}a_{31} = 0$ which requires (since $\gamma \neq$ square) either $a_{13} = a_{23} = 0$ or $a_{31} = a_{32} = 0$ both cases being excluded as they give a row or column of zeros. If

$$[M_9] : \left| \begin{pmatrix} x_1 & x_2 & 0 \\ 0 & x_1 & x_2 \\ 0 & 0 & 0 \end{pmatrix} + (a_{ij})x_3 \right| = 0$$

we may as before suppose $a_{22} = a_{23} = 0$. Then from the terms in $x_1{}^2x_3$, $x_2{}^2x_3$, and $x_1x_2x_3$ we get $a_{33} = a_{31} = a_{32} = 0$ and hence a zero row, which excludes the case. If

---

[25] H. R. Brahana, *American Journal of Mathematics*, vol. 56 (1934), pp. 490-510.

$$[M_{10}] : \left| \begin{pmatrix} x_1 & 0 & 0 \\ 0 & x_1 & x_2 \\ x_2 & 0 & 0 \end{pmatrix} + (a_{ij})x_3 \right| = 0,$$

we may suppose $a_{22} = a_{23} = 0$. Then the terms in $x_1^2x_3$ and $x_2^2x_3$ give $a_{33} = a_{12} = 0$, and the one in $x_1x_2x_3$ gives $a_{13} = -a_{32}$. Then $|a_{ij}| = -a_{13}^2a_{21} = 0$ gives $a_{21} = 0$ for otherwise $M(0, 0, 1)$ is of rank one. Now let $x'_3 = a_{13}x_3$, giving $a'_{13} = 1$. Then the terms in $x_1x_3^2$ and $x_2x_3^2$ give $a_{11} = a_{31} = 0$. Interchanging the first two rows gives $M^{(0)}(x)$ representing the only τ-class with no x-section of rank one.

A similar consideration of M_{11}, M_{12}, M_{13}, and M_{14} shows that $M^{(1)}$ and $M^{(2)}$ represent the only other τ-classes with $f(x) \equiv 0$ and actually belonging to the case $(3, 3, 3)$.

10. Uniqueness properties of τ-classes corresponding to groups $G(l, 3, 3)$. We have in sections 4-9 determined the τ-classes $(l, 3, 3)$, $l = 0, 1, 2, 3$ which derivation carries with it, as we have seen in section 3, the classifications for $l = 9, 8, 7, 6$. The case $l = 4$ would involve among other things a classification of quaternary cubic forms and will not be treated here.

In section 2, we saw that certain groups G define more than one τ-class. But a group G is completely defined by a single τ-class g. Hence the possibility of defining G by a second τ-class g must be property of the τ-class g itself, entirely aside from its relation to the group. This, indeed, is merely another way of saying that a collection E of forms is defined by any single form in it.

Suppose that the τ-class g defines G with generating subgroups S and U with the required properties (§ 1). In looking for generating subgroups S' and U' of G which might define another τ-class g' we need not consider subgroups obtained from S and U by means of operations under which the τ-class is invariant, viz. isomorphisms of G; new choices of generators in S, U, and K; and τ. The generators of S' and U' may be written in the form $r^xu^ys^z$ where we may, by means of isomorphisms of G, drop the r^x. Suppose now that

$$S' = \{u^{y^{(1)}}s^{z^{(1)}}, \cdots, u^{y^{(k')}}s^{z^{(k')}}\} \quad \text{and} \quad U' = \{u^{\eta^{(1)}}s^{\zeta^{(1)}}, \cdots, u^{\eta^{(m')}}s^{\zeta^{(m')}}\}$$

gives the τ-class g'. The changes induced on the variables and coefficients of the trilinear form will take a form in g into one in g'. The induced change will therefore be a linear transformation on the variables y and z taken together, and of such a nature that the derived form is still linear in the new sets of variables y' and z'. Then the problem stated in terms of forms alone is: (1) under what circumstances will transformations of the form:

$$y'_i = \sum_{j=1}^{m} a_{ij} y_j + \sum_{j=1}^{k} a_{i,j+m} z_j, \qquad (i = 1, 2, \cdots, m');$$

$$z'_i = \sum_{j=1}^{m} a_{i+m',j} y_j + \sum_{j=1}^{k} a_{i+m',j+m} z_j, \qquad (i = 1, 2, \cdots, k),$$

where (a_{ij}) is non-singular, replace a given trilinear form in variables x, y, z by another trilinear form in variables x, y', z'? (2) under what circumstances when (1) is possible will the derived form belong to a different τ-class from that of the initial form? We shall not attempt to answer these questions in general but will examine the τ-classes $(l, 3, 3)$ and answer question (2) for them. As an aid in this we introduce the concept of *separable* groups (and forms).

DEFINITION. *If generators of a group $G(l, m, k)$ can be chosen so that $r_{ij} = 1$, $(i > m_1, j \leq k_1)$ and $(i \leq m_1, j > k_1)$ we shall say that G is separable into the components*

$$G_1 = \{u_1, \cdots, u_{m_1}, s_1, \cdots, s_{k_1}\} \quad \textit{and} \quad G_2\{u_{m_1+1}, \cdots, u_m, s_{k_1+1}, \cdots, s_k\}.$$

The groups G_1 and G_2 are also groups "G" and may themselves be separable. If so we may continue the separation process until we have in G the *non-separable* or inseparable component groups $G_1, G_2, \cdots, G_\lambda$ no two of which have in common operators outside the commutator subgroup, and such that $G = \{G_1, G_2, \cdots, G_\lambda\}$. Such a separation will be called a *complete separation* of G.

The separable groups $(l, 3, 3)$ are equivalent to the following three:

$$G^{(1)} = \begin{pmatrix} r_{11} & 1 & 1 \\ 1 & r_{22} & 1 \\ 1 & 1 & r_{33} \end{pmatrix}, \quad G^{(2)} = \begin{pmatrix} r_{11} & 1 & 1 \\ 1 & r_{22} & r_{23} \\ 1 & r_{32} & r_{33} \end{pmatrix}, \quad G^{(3)} = \begin{pmatrix} 1 & r_{12} & r_{13} \\ r_{21} & 1 & 1 \\ r_{31} & 1 & 1 \end{pmatrix},$$

where the r_{ij} may or may not satisfy further relations.

Given $G = \{s_1, s_2, s_3, u_1, u_2, u_3\}$ and a τ-class g defined by this representation of G we ask if there can exist subgroups Σ and Γ in G satisfying our hypothesis for generating subgroups (§ 1) and such that the τ-class g' defined by the representation $G = \{\Sigma, \Gamma\}$ of G is different from g. This might be possible for some τ-classes and not for others. If Σ is of order $p^{m'}$ and Γ of order $p^{k'}$ we must have $m' + k' = 6$; for $m + k + l$ and l, and therefore

$m + k$ are invariants of G. We may express any generator of Σ or Γ as $r^x u^y s^z$ where, as we have seen, we lose no generality in suppressing the r^x.

First suppose that $\Sigma\{\sigma_1, \sigma_2, \sigma_3\}$ is of order p^3. Then $\Gamma = \{\gamma_1, \gamma_2, \gamma_3\}$ is likewise of order p^3. Let $\sigma_i = u^{y^{(i)}} s^{z^{(i)}}$; $\gamma_i = u^{\eta^{(i)}} s^{\zeta^{(i)}}$; $i = 1, 2, 3$. Now if the $s^{z^{(i)}}$ generate S we may write a) $G = \{\Sigma, U\}$ and the commutator structure will remain unaltered since $u^y s^z$ has the same commutator as s^z with $u^{y'}$. Similarly if: b) $G = \{\Gamma, U\}$, c) $G = \{S, \Sigma\}$ or d) $G = \{S, \Gamma\}$ there is no change in the defining relations. Any change in the τ-class then would have to come in the second step when the other of the new generating subgroups is introduced. Given $G = \{\Sigma, U\}$ we may express the γ_i in terms of the generators of Σ and of U, thus having the problem: given $G = \{S, U\}$ to find Γ so that $G = \{S, \Gamma\}$ etc. But then the $u^{\eta^{(i)}}$ must generate U and the τ-class remains unchanged under the replacement.

The only case then that we need consider is that in which none of a), b), c), d) above are possible, i. e. Σ and Γ each contain at least one operator in common with S and U. By new choice of generators in the four subgroups we may suppose $\Sigma = \{s_1, u^y s^z, u_3\}$, $\Gamma = \{u_1, u^{y'} s^{z'}, s_3\}$. We first require that G be non-separable. Now one of z_2 and z'_2 is different from zero. By at most change of notation, we may suppose $z_2 \neq 0$. Then a new choice of generators gives $s^z = s_2$. Now for Σ and Γ to be abelian we must have $r_{13} = r_{31} = r_{32} = 1$. If $y'_2 \neq 0$ we have also $r_{23} = 1$ and G is separable. Hence, we suppose $y'_2 = 0$. Then $z'_2 y_2 \neq 0$ since Σ and Γ generate G. But this gives $r_{12} = r_{21} = 1$ and G is again separable. Hence, *if G is non-separable it can belong to only one τ-class* $(l, 3, 3)$.

Next suppose Σ of order p^4. Now if $\{s^{z^{(i)}}\} = S$, some operator from U in Σ would be permutable with S contrary to hypothesis on G. Hence, we may choose generators in S, U, and Σ so that $\Sigma = \{s_1, s_2, u_1, u_2\}$; which requires $r_{11} = r_{12} = r_{21} = r_{22} = 1$. Now let $\Gamma = \{u^{y^{(1)}} s^{z^{(1)}}, u^{y^{(2)}} s^{z^{(2)}}\}$. Since $G = \{\Sigma, \Gamma\}$ we cannot have $\begin{vmatrix} y_3{}^{(1)} & y_3{}^{(2)} \\ z_3{}^{(1)} & z_3{}^{(2)} \end{vmatrix} = 0$ so we may suppose $u^{y^{(1)}} = u_3$ and then $y_3{}^{(2)} = 0$, $s_3{}^{(2)} \neq 0$ and by new choice of generators we may suppose $\Gamma = \{u_3 s^{z^{(1)}}, u^{y^{(2)}} s_3\}$ where $z_3{}^{(1)} = y_3{}^{(2)} = 0$. Hence $r_{33} = 1$ and G is separable and isomorphic to one of the groups $G^{(3)}$, listed above. Σ of order p^5 is evidently impossible for G separable or inseparable.

Now consider G separable. The groups $G^{(3)}$ as we have just seen can be defined by τ-classes $(l, 2, 4)$ and hence cannot be isomorphic to the groups $G^{(1)}$ and $G^{(2)}$ unless they can likewise be thus defined. But in neither $G^{(1)}$ nor $G^{(2)}$ are there pairs of permutable subgroups of order p^2 one each from S and U. Further the groups $G^{(3)}$ are the only ones which would be found among the groups $G(l, m, k)$ for two different solutions of $m + k = 6$. These groups were classified by Brahana.[26] Since there is at most one τ-class for each value of l, the groups $G^{(3)}$ present no uniqueness problem.

The groups $G^{(2)}$ and $G^{(1)}$ are readily distinguished by various properties (characteristic subgroup structure etc.) and so none of them can be defined by two different τ-classes. Our conclusion then is that except for the three groups $G^{(3)}$ for $l = 2, 3, 4$ each group $G(l, 3, 3)$ defines just one τ-class.

11. Deep lying nature of group theoretic properties derived from trilinear form invariants. For an interpretation of some of the form invariants in terms of the groups suppose that $r_1, \cdots, r_l$ generate K and that $u_i^{-1}s_j^{-1}u_is_j = \prod_h r_h^{a_{hij}}$. Now we ask conditions on z so that s^z will be permutable with some operator in U. For this it is necessary and sufficient that the commutators of s^z with $u_1, \cdots, u_m$ generate a group of order less than p^m. u_i and s^z have the commutator $\prod_h r_h^{\sum_j a_{hij}z_j}$. Hence, for these commutators to be related we must have z such that the matrix $M_z = (\sum_j a_{hij}z_j)$ with m rows and l columns be of rank less than m. For $l < m$ this is true for all z. For $l = m = k = 3$ we have as our condition $| M_z | = 0$, i. e. z is a point on the z-cubic related to the form. It will be permutable with u^y where y must be on the y-cubic. If M_z is of rank two s^z will be permutable with just one subgroup u^y. If M_z is of rank one s^z will be permutable with a subgroup of order p^2 in U. If no z-section is of rank less than three evidently no y-section could be of rank less than three, for if so some u^y would be permutable with s^z and hence M_z would be of rank less than three. But M_z is symmetric in x and y. Hence, if $| M_z | = 0$ is a cubic without real points, then $f(x) = | M_x | = 0$ is a cubic without real points and conversely. So for $m = k = l = 3$ we have as a necessary and sufficient condition that no operator in S be permutable with any operator in U that $f(x)$ be the "imaginary" cubic. (Class 20).

Proceeding similarly we could obtain numerous theorems of this character. However, from the complexity of the considerations involved it seems unlikely that we will be able to express in ordinary group theoretic terms the non-isomorphism of many of the groups which we have proved non-isomorphic by the above algebraic considerations. For instance, consider the two groups defined by the matrix

$$M_x(x) = \begin{pmatrix} x_1 & x_3 & 0 \\ 0 & x_1 & x_3 \\ 0 & 0 & x_1 \end{pmatrix} + x_2 \begin{pmatrix} 0 & 0 & 0 \\ a & 0 & 0 \\ 0 & -1-a & 0 \end{pmatrix}$$

---

[26] H. R. Brahana, *American Journal of Mathematics*, vol. 57 (1935), pp. 645-667.

for two values a_1 and a_2 of a, $a_i \neq 0, -1$; $a_1 + a_2 \neq -1$. We have proved that these groups G_1 and G_2 are non-isomorphic. Let us review the properties which these groups have in common. They are metabelian groups of the same order, p^9; both conformal with the abelian group of order p^9 and type $1, 1, 1, \cdots$; each contains maximal invariant subgroups of order p^6 which with an abelian subgroup of its group of isomorphisms will generate the whole group. Generators of G_1 and G_2 can be so chosen that the totality of operators s^x permutable with operators in U is the same in each case and also the totality of operators u^y permutable with operators in S is the same in each case. In both groups every commutator different from identity is of the first kind and extent two. Furthermore the subgroups are the same in the sense that if we consider all of the subgroups of G_1 isomorphic to a given one there will be the same number of subgroups in G_2 isomorphic to the named one. The quotient groups are similarly related. The difference between the two groups lies in the different arrangements relative to each other of the subgroups.

This situation is somewhat analogous to the projective classification of sets of five (no three collinear) points in the plane. Two such sets may be projectively non-equivalent and yet the subsets of the two are projectively equivalent.

A complete classification of the groups G then from a group theoretic standpoint must involve the introduction of new group theoretic conceptions, perhaps an extension of the "type" and "extent" defined above.

UNIVERSITY OF ILLINOIS.

BIHARMONIC FUNCTIONS IN ABSTRACT SPACES.*

By A. E. Taylor.

1. Introduction. In classical analysis there is a remarkable parallelism between the theory of analytic functions of a complex variable and the theory of harmonic functions of two real variables; the source of this relationship lies in the Cauchy-Riemann equations. When we consider analytic functions of n complex variables there is a corresponding parallelism with a class of functions of $2n$ real variables. This class is usually designated by the name biharmonic; it is much more restricted than the class of solutions of Laplace's equation in $2n$ dimensions. In fact, a biharmonic function of $x_1, \cdots, x_n, y_1, \cdots, y_n$ satisfies the system of n^2 equations.[1]

$$(1) \qquad \frac{\partial^2 u}{\partial x_\mu \partial x_\nu} + \frac{\partial^2 u}{\partial y_\mu \partial y_\nu} = 0 \qquad \frac{\partial^2 u}{\partial x_\mu \partial y_\nu} - \frac{\partial^2 u}{\partial x_\nu \partial y_\mu} = 0.$$

It has been shown elsewhere that the Cauchy-Riemann equations may be generalized to meet the needs of the theory of analytic functions on one normed vector space to another.[2] It is the purpose of this note to make the natural extension of the notion of a biharmonic function which is suggested by the generalization just mentioned, and to prove a fundamental theorem pertaining to such functions. As an illustration we consider real-valued biharmonic functions of the doubly infinite set $(x_1, x_2, \cdots; y_1, y_2, \cdots)$ where each of the sequences $x = \{x_n\}$, $y = \{y_n\}$ is taken to be a point of the Hilbert space (l_2), that is, $\sum_1^\infty |x_n|^2$ is finite, and similarly for $\{y_n\}$.

2. The Cauchy-Riemann equations. Let E, E' be real, Banach spaces. With E we can associate a complex space $E(C)$ of couples of elements from E. If (x, y) is such a couple, and a, b are real numbers, we define

$$(x_1, y_1) + (x_2, y_2) = (x_1 + x_2, y_1 + y_2)$$
$$(a + ib) \cdot (x, y) = (ax - by, bx + ay)$$
$$\| (x, y) \| = (\| x \|^2 + \| y \|^2)^{\frac{1}{2}}.$$

---

* Received September 28, 1937.

[1] W. F. Osgood, *Lehrbuch der Funktionentheorie*, vol. II, 1 (1929), pp. 22-23.

[2] A. E. Taylor, "Analytic functions in general analysis," *Annali della Reale Scuola Normale Superiore di Pisa*, (in press) § 8. We shall refer to this as paper (A). See also *Comptes Rendus*, vol. 203 (1936), pp. 1228-1230.

Two couples are regarded as equal if and only if their corresponding members are equal. If we write x for $(x, 0)$ we may also write $(x, y) = x + iy$, as with complex numbers. We shall denote $x + iy$ by the single letter z. Clearly $E(C)$ is also a Banach space (complex). In a similar fashion we construct $E'(C)$.

We may also form a *real* space of these same couples, defining the product $a(x, y)$ only when a is real: $a(x, y) = (ax, ay)$, and norming it in the same way. This space, which we denote by E^2, is likewise complete; as a metric space it is indistinguishable from $E(C)$.

Let $f(z) = f_1(x, y) + if_2(x, y)$ be a function defined on an open set D of $E(C)$, its values being in $E'(C)$. *It is said to be analytic in D if it is continuous there and possesses a variation at each point of D.*[3] If $f(z)$ is analytic the essential properties of the functions f_1, f_2 may be stated as follows (Theorem 17, paper (A)).

THEOREM 1. *In order that $f(z)$ be analytic in D it is necessary and sufficient that f_1, f_2 be continuous and admit continuous first partial variations satisfying the equations*

$$\begin{aligned} \delta_\xi{}^x f_1(x, y) &= \delta_\xi{}^y f_2(x, y) \\ \delta_\xi{}^y f_1(x, y) &= -\delta_\xi{}^x f_2(x, y) \end{aligned} \tag{2}$$

at all points of D, for an arbitrary element ξ of E. Both f_1 and f_2 then admit continuous total Fréchet differentials of all orders in D.[4]

3. Biharmonic functions. Let us first recall an important symmetry property of Fréchet differentials: if $F(x)$ is a function on one complete normed vector space to another which is defined in the neighborhood of x_0, and possesses continuous first and second Fréchet differentials there, then $d_{y_2}{}^x d_{y_1}{}^x F(x) = d_{y_1}{}^x d_{y_2}{}^x F(x)$ at each point of the neighborhood.[5] If we apply this to the functions f_1, f_2 of § 2 we find symmetry relations of the type

---

[3] A function $F(x)$ is said to have a variation at x_0 if it is defined in the neighborhood of x_0 and if $\delta_y{}^x F(x_0) = \lim\limits_{t \to 0} \dfrac{F(x_0 + ty)}{t}$ exists for every y. The variable t is real or complex according as the spaces are real or complex.

[4] By the *total* differential of $f_1(x, y)$ is meant the differential with respect to the composite variable (x, y), an element of the space E^2. The partial Fréchet differentials of $f_1(x, y)$ exist also, and coincide with the partial variations, so that the total differential is

$$d_{(\xi,\eta)}{}^{(x,y)} f_1(x, y) = \delta_\xi{}^x f_1(x, y) + \delta_\eta{}^y f_1(x, y).$$

See T. H. Hildebrandt and L. M. Graves, *Transactions of the American Mathematical Society*, vol. 29 (1927), pp. 136-138. We use the Latin d for Fréchet differentials.

[5] M. Kerner, *Annals of Mathematics*, vol. 34 (1933), p. 549.

$$d_\eta^x d_\xi^x f_1(x, y) = d_\xi^x d_\eta^x f_1(x, y), \qquad d_\eta^x d_\xi^y f_1(x, y) = d_\xi^y d_\eta^x f_1(x, y)$$

and certain others which need not be written down because of their similarity. Making use of these, we find upon differentiating equations (2) that the functions f_1, f_2 must satisfy the conditions

$$(3) \qquad \begin{aligned} d_\eta^x d_\xi^x f_k(x, y) + d_\eta^y d_\xi^y f_k(x, y) &= 0 \\ d_\eta^x d_\xi^y f_k(x, y) - d_\eta^y d_\xi^x f_k(x, y) &= 0 \end{aligned} \qquad (k = 1, 2)$$

at each point of D, for each element (ξ, η) of E^2. These equations are evidently a generalization of (1), and are equivalent to the latter if E and E' coincide with the real Euclidean n-space. Accordingly we lay down the following definition:

Definition. A function $u(x, y)$ which is defined in an open set D of E^2, with values in E', is said to be biharmonic in D if it is continuous in D and possesses first and second total Fréchet differentials which are continuous in D and there satisfy equations (3).

In order to be assured of the appropriateness of this terminology it is necessary to show that a biharmonic function is derivable from a suitable analytic function on $E(C)$ to $E'(C)$, of which it is the 'real' part. We shall do this. First we must consider an existence theorem for what may be called 'exact differentials.'

THEOREM 2. *Let D be a simply connected*[6] *open set in the space E^2. Let $P(x, y, \xi)$, $Q(x, y, \xi)$ be functions with values in E', defined when (x, y) is in D and ξ is in E; furthermore let P and Q be linear in ξ, and possess continuous Fréchet differentials with respect to (x, y) such that for (x, y) in D, ξ_1, ξ_2 in E*

$$(4) \qquad \begin{aligned} d_{\xi_2}^x P(x, y, \xi_1) &= d_{\xi_1}^x P(x, y, \xi_2) \\ d_{\xi_2}^y Q(x, y, \xi_1) &= d_{\xi_1}^y Q(x, y, \xi_2) \\ d_{\xi_2}^y P(x, y, \xi_1) &= d_{\xi_1}^x Q(x, y, \xi_2). \end{aligned}$$

Then there exists a function $F(x, y)$ on D to E', unique apart from an additive constant, whose total Fréchet differential exists and is precisely $P(x, y, \xi) + Q(x, y, \eta)$.

This theorem is an immediate consequence of a theorem of Kerner.[7] By means of it we now readily establish our principal result:

---

[6] We use the definition given by Kerner, *loc. cit.*, p. 555.

[7] M. Kerner, *loc. cit.*, p. 555, Theorem 3.

THEOREM 3. *Let* $u(x,y)$ *be biharmonic in a simply connected open set* D *of the space* E^2. *Then there exists a second function* $v(x,y)$ *which is also biharmonic in* D, *and such that the couple* $(u,v) = u + iv$, *regarded as a function on* $E(C)$ *to* $E'(C)$, *is analytic in* D. *The function* $v(x,y)$ *is unique apart from an additive constant.*

Proof. Consider the functions

$$P(x,y,\xi) = -d_\xi{}^y u(x,y),\ Q(x,y,\eta) = d_\eta{}^x u(x,y).$$

They satisfy the hypotheses of Theorem 2 by virtue of (3) and the properties of Fréchet differentials. Therefore there is defined in D a function $v(x,y)$ with the continuous differential $-d_\xi{}^y u(x,y) + d_\eta{}^x u(x,y)$. But also, the differential of $v(x,y)$ is $d_\xi{}^x v(x,y) + d_\eta{}^y v(x,y)$. Hence u and v are continuous, and together satisfy the Cauchy-Riemann equations (2); therefore $u + iv$ is analytic in D. From this it follows that v is also biharmonic.

4. Biharmonic functions of an infinite number of variables. Let E be the Hilbert space (l_2) and E' the real number system. Then $E(C)$ is the complex Hilbert space H_0 analagous to (l_2), and E^2 is the space of couples (x,y), where $x = (x_1, x_2, \cdots)$, $y = (y_1, y_2, \cdots)$. The typical element of H_0 is $z = x + iy = (x_1 + iy_1, x_2 + iy_2, \cdots)$. Let D be an open set in H_0 (we may also regard it as an open set in $(l_2)^2$). If $f(z) = f_1(x,y) + if_2(x,y)$ is a complex function defined on D the conditions for analyticity may be stated as follows.

THEOREM 4. *In order that* $f(z)$ *be analytic in* D *it is necessary and sufficient that* $f(z)$ *be continuous in* D *and possess first partial derivatives* $\dfrac{\partial f}{\partial z_\nu}$ $(\nu = 1, 2, \cdots)$ *at each point of* D. *Stated in terms of* f_1 *and* f_2 *the conditions are:* f_1 *and* f_2 *shall be continuous in* D *and possess continuous first partial derivatives with respect to each of the real variables* x_ν, y_ν, and in *addition, the equations*

$$(5)\qquad \frac{\partial f_1}{\partial x_\nu} = \frac{\partial f_2}{\partial y_\nu} \qquad \frac{\partial f_1}{\partial y_\nu} = -\frac{\partial f_2}{\partial x_\nu} \qquad (\nu = 1, 2, \cdots)$$

must be satisfied.

Proof. We consider the conditions on $f(z)$ first. They are clearly necessary. They are also sufficient. It is enough to prove this for an open set D of the type $\| z \| < r$. Corresponding to each z of this set and each positive integer n we consider the function $f_n(z) = f(z^{(n)})$, where

$$z^{(n)} = (z_1, z_2, \cdots, z_n, 0, \cdots).$$

Since $\| z^{(n)} \| \leq \| z \|$ these functions are continuous in D, and hence analytic, for it is easily seen that they possess the variations

$$\delta_w{}^z f_n(z) = \sum_{\nu=1}^{n} f_{z_\nu}(z^{(n)}) w_\nu \qquad \left(f_{z_\nu} = \frac{\partial f}{\partial z_\nu} \right).$$

The analyticity of $f(z)$ in D is then a consequence of the fact that as n tends to infinity, $f_n(z) \rightarrow f(z)$, the convergence being uniform in each compact set extracted from the closed sphere $\| z \| \leq \theta r$ $(0 < \theta < 1)$. For let G be such a compact set, and let $\epsilon > 0$ be given. Denote by G' the set consisting of all points of G together with all points $z^{(n)}$ $(n = 1, 2, \cdots)$ where z is in G. Then it is not difficult to see that G' is compact and lies in the same closed sphere with G. Now $f(z)$ is uniformly continuous in the compact set G' (paper (A), Theorem 4). Let us choose δ so that $\| z' - z \| < \delta$ implies $| f(z') - f(z) | < \epsilon$ whenever z, z' are in G'. Fréchet has shown[8] that corresponding to the compact set G' there exists a convergent series $\sum_1^\infty a_\nu{}^2$ such that $\sum_{\nu=n-1}^{\infty} | z_\nu |^2 < \sum_{\nu=n+1}^{\infty} a_\nu{}^2$ for all n and all z in G'. Hence if we choose N so that $n \geq N$ implies $\sum_{\nu=n+1}^{\infty} a_\nu{}^2 < \delta^2$ we shall have $\| z^{(n)} - z \|^2 = \sum_{\nu=n+1}^{\infty} | z_\nu |^2 < \delta^2$, and $| f(z) - f_n(z) | < \epsilon$ for all z in G whenever $n \geq N$. $f(z)$ is then analytic, by an extension of a theorem of Weierstrass (Theorem 13 of paper (A)). Since it is known that the variation of an analytic function is analytic and is, in fact, its Fréchet differential, we are able to infer that the partial derivatives of $f(z)$ are analytic, and that $d_w{}^z f(z) = \sum_{\nu=1}^{\infty} \frac{\partial f}{\partial z_\nu} w_\nu$. The remainder of the theorem is now deducible by classical methods.

Finally we consider conditions analogous to (1) which will assure us that a real function $u(x, y)$ defined on an open set D in $(l_2)^2$ is biharmonic there.

THEOREM 5. *Let $u(x, y)$ be defined and continuous in D, and possess continuous first and second partial derivatives with respect to the variables x_ν, y_ν. Let*

$$a_{\mu\nu} = \frac{\partial^2 u}{\partial x_\mu \partial x_\nu}, \quad b_{\mu\nu} = \frac{\partial^2 u}{\partial x_\mu \partial y_\nu}, \quad c_{\mu\nu} = \frac{\partial^2 u}{\partial y_\mu \partial y_\nu}$$

and suppose that:

---

[8] M. Fréchet, *Rendiconti dell Circolo Matematico di Palermo*, vol. 30 (1910), pp. 18-19.

(i) *the series* $\sum_{\nu=1}^{\infty}\left[\left(\frac{\partial u}{\partial x_\nu}\right)^2+\left(\frac{\partial u}{\partial y_\nu}\right)^2\right]$ *converges uniformly in each compact set contained in an arbitrary closed sphere in* D;[9]

(ii) *each of the series* $\sum_{\nu=1}^{\infty} a^2_{\mu\nu}$, $\sum_{\nu=1}^{\infty} b^2_{\mu\nu}$, $\sum_{\nu=1}^{\infty} c^2_{\mu\nu}$ $(\mu=1,2,\cdots)$ *converges according to condition* (*c*) *in* D;

(iii) *for each* ξ *in* (l_2) *each of the sequences*

$$\sum_{\mu=1}^{n}\left|\sum_{\nu=1}^{\infty} a_{\mu\nu}\xi_\nu\right|^2,\qquad \sum_{\mu=1}^{n}\left|\sum_{\nu=1}^{\infty} b_{\mu\nu}\xi_\nu\right|^2,\qquad \sum_{\mu=1}^{n}\left|\sum_{\nu=1}^{\infty} c_{\mu\nu}\xi_\nu\right|^2$$

converges according to condition (*c*) *in* D;

(iv) *the equations*

$$a_{\mu\nu}+c_{\mu\nu}=0 \qquad b_{\mu\nu}-b_{\nu\mu}=0$$

are satisfied at each point of D (*compare with* (1) *of* § 1). *Then:* u *has continuous first and second total Fréchet differentials satisfying* (3) *of* § 3 *in* D, *and hence is biharmonic.*

Proof. We can at once infer, by methods due to Hart,[10] that the first differential exists, is continuous in D, and given by

$$d_{(\xi,\eta)}{}^{(x,y)}u(x,y)=\sum_{\nu=1}^{\infty}\left(\frac{\partial u}{\partial x_\nu}\xi_\nu+\frac{\partial u}{\partial y_\nu}\eta_\nu\right). \tag{6}$$

This all follows as a consequence of the hypotheses on u and its first partial derivatives. The next thing is to prove that the function

$$\phi(x,y,\xi,\eta)=d_{(\xi,\eta)}{}^{(x,y)}u(x,y),$$

when ξ,η are fixed, has these same properties. The existence and continuity of the partial derivatives of ϕ in D follows by classical methods, because of (ii) and the symmetry of the matrices $\{a_{\mu\nu}\}$, $\{b_{\mu\nu}\}$ and $\{c_{\mu\nu}\}$; $\{b_{\mu\nu}\}$ is symmetric, by (iv). We have

---

[9] We shall then, for brevity, say that the series converges according to condition (c) in D.

[10] W. L. Hart, *Transactions of the American Mathematical Society*, vol. 23 (1922), pp. 30-39. Our assumptions, though somewhat weaker than Hart's are equally effective, for by them we are enabled to show that the series in condition (i) defines a function which is bounded in each of the compact sets in question, and continuous in D; similarly for the function in (6) when ξ,η are fixed.

$$\frac{\partial\phi}{\partial x_\mu}=\sum_{\nu=1}^{\infty}a_{\mu\nu}\xi_\nu+\sum_{\nu=1}^{\infty}b_{\mu\nu}\eta_\nu$$
$$\frac{\partial\phi}{\partial y_\mu}=\sum_{\nu=1}^{\infty}b_{\nu\mu}\xi_\nu+\sum_{\nu=1}^{\infty}c_{\mu\nu}\eta_\nu\ ,$$

from which it follows rather easily by (iii) that the series $\sum_{\mu=1}^{\infty}\left[\left(\frac{\partial\phi}{\partial x_\nu}\right)^2+\left(\frac{\partial\phi}{\partial y_\mu}\right)^2\right]$ converges according to condition (c) in D, for each ξ. From (iii) it also follows, by a theorem of Hellinger and Toeplitz,[11] that the matrices $\{a_{\mu\nu}\}$, $\{b_{\mu\nu}\}$, $\{c_{\mu\nu}\}$ define bilinear forms which are, at each point of D, bounded in the sense of Hilbert. By repetition of the reasoning already referred to in obtaining the first differential of $u(x,y)$ we see that

$$d_{(\xi',\eta')}{}^{(x,y)}\phi(x,y,\xi,\eta)=\sum_{\mu,\nu=1}^{\infty}a_{\mu\nu}\xi'_\mu\xi_\nu$$
$$+\sum_{\mu,\nu=1}^{\infty}b_{\mu\nu}\xi'_\mu\eta_\nu+\sum_{\mu,\nu=1}^{\infty}b_{\nu\mu}\eta'_\mu\xi_\nu+\sum_{\mu,\nu=1}^{\infty}c_{\mu\nu}\eta'_\mu\eta_\nu .$$

The remainder of the theorem then follows at once by use of condition (iv) and the definition of biharmonicity.

CALIFORNIA INSTITUTE OF TECHNOLOGY.

---

[11] *Mathematische Annalen*, vol. 69 (1910), pp. 321-322. See also M. H. Stone and J. D. Tamarkin, *Duke Mathematical Journal*, vol. 3 (1937), p. 298.

NORMAL COORDINATES FOR EXTREMALS TRANSVERSAL TO A MANIFOLD.*

By Stewart S. Cairns.

1. Introduction. Consider a positive definite regular calculus of variations problem defined on an n-dimensional manifold R throughout a neighborhood of a point q_0. A normal coördinate system[1] with origin q_0 is a system (y) in terms of which the extremals with q_0 for initial point can be represented, near the initial point, by linear equations $y^i = \lambda_i s$ $(0 \leq s < s_0)$.

In the present paper, we first obtain normal coördinates under hypotheses weaker than those heretofore used. We then define a new kind of coördinates, $(z) \equiv (z^1, \cdots, z^n)$, called *normal coördinates with respect to M near q_0*, where M is an m-dimensional manifold on R passing through q_0. In terms of (z), M is defined near q_0 by the equations $z^{m+1} = \cdots = z^n = 0$, and the general extremal cut transversally by M at its initial point, near q_0, is given by $z^i = a_i$ $(i = 1, \cdots, m)$, $z^j = \lambda_j s$ $(j = m + 1, \cdots, n)$ $(0 \leq s < s_0)$ where s represents arc length and where $\sum_{j=m+1}^{n} \lambda_j \lambda_j = 1$.

Underlying much of our work is a study by Morse[2] of the extremals cut transversally at their initial points by a manifold M. The subset of these extremals consisting of all those with a given initial point q covers an $(n - m)$-manifold which is differentiable near q save, in general, for a conical point at q. We obtain necessary and sufficient conditions on our calculus of variations problem that this manifold be differentiable even at q, independently of the particular manifold M.

---

* Received June 18, 1937; Revised July 26, 1937. This paper was written while the author was on leave of absence from Lehigh University and was a member of the Institute for Advanced Study at Princeton, N. J.

[1] An existence proof for normal coördinates is partly given in Duschek-Mayer, *Lehrbuch der Differentialgeometrie*, vol. 2, ch. V, §§ 5, 6, and is completed by J. H. C. Whitehead, "On the covering of a complete space by the geodesics through a point (§ 2)," *Annals of Mathematics*, vol. 36 (1935), pp. 679-704. The original relevant investigations both of the extremals with given initial point and of transversal extremals were made, in the case where R is euclidean 3-space, by Bliss and Mason, "The properties of curves in space which minimize a definite integral," *Transactions of the American Mathematical Society*, vol. 9 (1908), pp. 440-466; "Fields of extremals in space," *Transactions of the American Mathematical Society*, vol. 11 (1910), pp. 325-340.

[2] Marston Morse, "The calculus of variations in the large," *American Mathematical Society Colloquium Publications*, vol. 18 (1934), p. 111.

2. Definition of the metric. By an *n-manifold* $(n = 1, 2, \cdots)$, we mean a connected topological space which can be covered by a denumerable set of neighborhoods, each the homeomorph of an open region of euclidean n-space. Let coördinate systems $(x), (y), \cdots$ be introduced, by homeomorphisms, on the neighborhoods of such a set. The manifold is said to be *of class* C^ν in terms of the coördinate systems $(x), (y), \cdots$, which are called *admissible systems*, if every transformation between two of the systems is given by functions $x^i(y)$ of class[3] C^ν with a non-vanishing jacobian. Any further coördinate system is *admitted* if the transformations between it and the original systems are all of class C^ν with non-vanishing jacobians.

We suppose that R is an n-manifold $(n > 1)$ of class C^2. Let q_0 be a point on R and let (x_0) denote the coördinates of q_0 in some admissible system (x). Consider a calculus of variations problem whose basic function,[4] $F(x, r)$, is defined for (x) in some neighborhood of (x_0) and for $(r) \neq 0$. We suppose that F and F_{r^i} are of class C^1, that F is positive homogeneous of the first order in (r), and that

$$F(x, r) > 0, \qquad F_1(x, r) \neq 0 \qquad \text{when } (r) \neq 0, \tag{2.1}$$

where F_1 is defined (Morse, *op. cit.*, p. 112) by the identity

$$\begin{vmatrix} F_{r^i r^j} & u_i \\ v_j & 0 \end{vmatrix} \equiv -F_1(x, r)\,[r^i u_i]\,[r^j v_j]. \tag{2.2}$$

Our variations problem defines a metric, $ds = F(x, dx)$, *on any region of* R *throughout which the above conditions are fulfilled.*

3. Normal coördinates with origin q_0. By a *unit* contravariant vector (ρ) at (x), we mean one satisfying the equation

$$F(x, \rho) = 1. \tag{3.1}$$

There is a unique extremal tangent at a given initial point (x_0) to a given vector (r_0). We fix the parameter, t, on this extremal by requiring that $F(x, \dot{x}) = F(x_0, r_0)$ and that $x^i(0) = x_0{}^i$. The extremal, with its parameter, is then determined by the following system of equations, in which (ρ_0) denotes the unit vector in the same direction as (r_0):

$$\begin{aligned} &\frac{d}{dt} F_{r^i}(x, \dot{x}) - F_{x^i}(x, \dot{x}) = 0, \\ &x^i(0) = x_0{}^i, \qquad \dot{x}^i(0) = r_0{}^i = \tau\rho_0{}^i, \\ &F(x, \dot{x}) = F(x_0, \tau\rho_0) = \tau. \end{aligned} \tag{3.2}$$

---

[3] That is, continuous along with all derivatives of orders $\leq \nu$.

[4] For the laws of transformation of this function, see Morse, *loc. cit.*

The parameter t is related to the arc length, s, measured from (x_0) by the identity

$$s = \tau t = F(x_0, r_0)t. \tag{3.3}$$

Following a method to be found in Morse [*op. cit.*, Ch. V, § 4], we next consider the system

$$\frac{d}{dt}[(1+\lambda)F_{r^i}(x,\dot{x})] - F_{x^i}(x,\dot{x}) = 0, \quad F(x,\dot{x}) = \tau, \quad (\tau > 0), \tag{3.4}$$

where τ is a constant. If, in a solution $[x^i(t), \lambda(t)]$, we have $\lambda(0) = 0$, then $\lambda(t)$ is identically zero. [For proof, see Morse, *loc. cit.*].

When $F = \tau$, we have

$$\begin{vmatrix} F_{r^i r^j} & F_{r^i} \\ F_{r^j} & 0 \end{vmatrix} = -\tau^2 F_1 \neq 0. \tag{3.5}$$

It is therefore possible, by the implicit function theorem, to solve the equations

$$(1+\lambda)F_{r^i}(x,r) = v^i, \qquad F(x,r) = \tau \tag{3.6}$$

for $(r^1, \cdots, r^n, \lambda)$ in a neighborhood of $\lambda = 0$. The solutions,

$$r^i = r^i(x,v,\tau) = \tau r^i(x,v,1), \qquad \lambda = \lambda(x,v,\tau) \tag{3.7}$$

are of class C^1.

In place of the system (3.2), we shall use the system

$$\begin{aligned} \frac{dx^i}{dt} &= r^i(x,v,\sigma) = \sigma r^i(x,v,1), \\ \frac{dv_i}{dt} &= q_i(x,v,\sigma) = F_{x^i}(x, r(x,v,\sigma)) = \sigma q_i(x,v,1), \\ x^i(0) &= x_0{}^i, \\ v_i(0) &= v_{i0} = F_{r^i}(x_0, r_0), \end{aligned} \tag{3.8}$$

where σ is a positive constant. We write the general solution for x^i in the system (3.8) as follows:

$$x^i = h^i(t, \sigma, x_0, v_0). \tag{3.9}$$

In the case where $\sigma = \tau = F(x_0, r_0)$, the system (3.8) is equivalent to (3.4) with the initial conditions $[x^i(0) = x_0{}^i, \lambda(0) = 0]$ and is therefore equivalent to (3.2), since $\lambda(0) = 0$ implies that $\lambda(t)$ is identically zero. Hence, the extremals with q_0 for initial point are given by

$$\begin{aligned} x^i &= h^i(t, \tau, x_0, v_0) \\ &= h^i[t, F(x_0, r_0), x_0, F_{r^i}(x_0, r_0)] \equiv \phi^i(t, x_0, r_0). \end{aligned} \tag{3.10}$$

In view of identity (3.3), these equations can be written in the form

$$\begin{aligned} x^i &= \phi^i\left(\frac{s}{\tau}, x_0, \rho_0\tau\right) \\ &= \phi^i\left(\frac{s}{F(x_0, r_0)}, x_0, r_0\right) \equiv \psi^i(s, x_0, r_0). \end{aligned} \tag{3.11}$$

The functions ψ^i are solutions for (x) in the system (3.8) read for the special case where (r_0) is a unit vector and $\sigma = 1$.

(A) *The above method reveals the class* C^1 *character of* (ϕ^i, ψ^i) *and* $(\phi_t{}^i, \psi_s{}^i)$ *near* (x_0), *since the functions* r^i *in* (3.7) *and* (3.8) *are of class* C^1. *We note the following properties:*

$$\begin{aligned} &\phi^i(0, x_0, r_0) = \psi^i(0, x_0, r_0) = x_0{}^i, \\ &\phi_t{}^i(0, x_0, r_0) = r_0{}^i, \\ &\psi_s{}^i(0, x_0, r_0) = \rho_0{}^i. \end{aligned} \tag{3.12}$$

Let $s_0 > 0$ be so small that the representation (3.11) is valid $0 \leq s < s_0$ for all (ρ_0). The representation (3.10) is then valid $0 \leq t < s_0/\tau$. With (x_0, ρ_0) held fast, τ and r_0 can be so restricted in (3.10) and (3.11) that $\tau = F(x_0, r_0)$ satisfies the condition $0 \leq \tau < s_0$. This will result in no loss of generality in the results we have in mind. Under this restriction, the point where $t = 1$ is on the domain of (3.10). Since it coincides with the point where $s = \tau$, we have the identity

$$\psi^i(s, x_0, r_0) \equiv \phi^i(1, x_0, \rho_0 s) \qquad (0 < s < s_0), \tag{3.13}$$

which holds on each extremal with (x_0) for initial point. Hence, if we make the definition $\phi^i(1, x_0, 0) = x_0{}^i$, we can write the equations of all the extremals with q_0, for initial point in the form

$$x^i = \phi^i(1, x_0, \rho_0 s) \equiv g^i(x_0, \rho_0 s) \qquad (0 \leq s < s_0). \tag{3.14}$$

We now [cf. Mayer, *loc. cit.*] define the *normal coördinates* (y) with origin q_0 by the transformation

$$x^i = g^i(x_0, y). \tag{3.15}$$

This transformation is continuous throughout some neighborhood of (x_0). With the possible exception of the point (x_0), the partial derivatives exist and are continuous throughout such a neighborhood. The existence of these derivatives at (x_0) with the values

$$\left.\frac{\partial g^i}{\partial y^j}\right|_{(y)=(0)} = \delta_j^{\,i} \tag{3.16}$$

is established as follows. Let $\rho_{(j)}$ denote either of the unit vectors at q_0 with all (x)-components equal to zero save the j-th. Then at (x_0)

$$\begin{aligned}\lim_{\Delta y^j\to 0}\frac{\Delta_{y^j} g^i}{\Delta y^j} &= \lim_{s\to 0}\frac{g^i(x_0, s\rho_{(j)}) - g^i(x_0, 0)}{s\rho^j_{(j)}}\\ &= \frac{1}{\rho^j_{(j)}}\lim_{s\to 0}\frac{dg^i}{ds} = \frac{\rho^i_{(j)}}{\rho^j_{(j)}} = \delta_j^{\,i}.\end{aligned} \tag{3.17}$$

To establish the continuity of $\partial g^i/\partial y^j$ at $(y) = (0)$, we first note that

$$\frac{dg^i}{ds} = \rho_0^{\,j}\left(\frac{\partial g^i}{\partial y^j}\right), \tag{3.18}$$

since equation (3.15) is obtained from (3.13) by the substitution $(y) = (\rho_0 s)$. But dg^i/ds is the same as $\psi_s{}^i(s, x_0, r_0)$. Hence

$$\rho_0^{\,j}\left(\frac{\partial g^i}{\partial y^j}\right) = \psi_s{}^i(s, x_0, r_0). \tag{3.19}$$

The continuity of $\partial g^i/\partial y^j$ at $(y) = (0)$ follows readily from (A) together with equations (3.16) and (3.19). Furthermore, the jacobian $|\,\partial g^i/\partial y^j\,|$ has the value unity when $(y) = (0)$.

(B) *The transformation* (3.15) *to normal coördinates is therefore of class* C^1 *with a non-vanishing jacobian in some neighborhood of* $(y) = (0)$. *In terms of* (y), *the extremals with* q_0 *for initial point are given, near* q_0, *by* $y^i = \rho^i s$ $(0 \leqq s < s_0)$, *where* s *is arc length and* (ρ) *ranges over all the unit vectors at* q_0.

4. Transversal vectors. Let M be an m-manifold $(0 < m < n)$ of class C^2 on R. This means that M is intrinsically of class C^2, and also that, if $(x) = (x^1, \cdots, x^n)$ and $(\alpha) = (\alpha^1, \cdots, \alpha^m)$ are admissible coördinate systems on R and M respectively, having some point q_0 common to their domains, then, in a neighborhood of q_0 on M, we have

$$x^i = \phi^i(\alpha) \qquad (i = 1, \cdots, n), \tag{4.1}$$

where the ϕ's are of class C^2 and the functional matrix $|\,\partial\phi^i/\partial\alpha^j\,| = |\,\phi_j{}^i(\alpha)\,|$ is of rank m.

If (r^*) and (r) denote contravariant vectors at (x), then we say that (r^*) is *transversal* to (r) if the invariant equation

$$F_{r^i}(x, r)\, r^{*i} = 0 \tag{4.2}$$

holds. If (x) is on M and every vector (r^*) tangent to M at (x) is transversal to (r), we say that M is *transversal* to (r) at (x). The condition that M be transversal to (r) at (x) can be expressed in terms of the ϕ's as follows:

$$F_{r^i}[\phi(\alpha), r]\phi_h{}^i(\alpha) = 0 \qquad (h = 1, \cdots, m). \tag{4.3}$$

The solutions (r) of (4.3) will be called the *transversal vectors at* (α).

In the space of the vectors (r), the equation

$$F(x, r) = 1 \tag{4.4}$$

represents a closed convex[5] $(n-1)$-dimensional manifold of class C^2, called the *indicatrix* of our calculus of variations problem in the point (x). The following statement gives a geometric interpretation of transversality.

(A) Let (ρ) be a unit vector at $(x) \equiv [\phi(\alpha)]$, and let S_{n-1} be the indicatrix in the point (x). Let (τ_m) denote the m-plane tangent to M at (α). Then (ρ) is a unit transversal vector at (α) if and only if the $(n-1)$-plane tangent to S_{n-1} at (ρ), where (ρ) is now interpreted as a point on S_{n-1}, contains an m-plane parallel to τ_m.

(B) Let 0 be the origin in the space of the vectors (r) at (x) and let $S_{n-m-1}(\tau_m)$ be the set of all points on S_{n-1} at each of which the tangent $(n-1)$-plane to S_{n-1} contains an m-plane parallel to τ_m. Then the rays from 0 through $S_{n-m-1}(\tau_m)$ form a cone, referred to hereafter as the *transversal cone,* whose elements give the directions of the transversal vectors at (α).

5. The polar coördinates (α, λ, s). We next establish the following result.

THEOREM. *The transversal cone is the homeomorph of an $(n-m)$-plane.*

Proof. We first define *unit covariant vectors* at q near q_0 with reference to a set $\pi_i{}^{(j)}(q)$ $(j = 1, \cdots, n)$ of n covariant vectors, arbitrarily selected subject to the restrictions (1) that they be independent at each point q in some neighborhood of q_0 and (2) that their components be class C^1 functions of the (x)-coördinates of q. A further restriction will be imposed later. By a *unit covariant vector* π_i at q, we will mean any vector of the form

$$\pi_i = \lambda_j \pi_i{}^{(j)}(q) \qquad (i = 1, \cdots, n), \tag{5.1'}$$

where

$$\lambda_j \lambda_j = 1. \tag{5.1''}$$

---

[5] Cf. C. Carathéodory, *Variationsrechnung* (1935), §§ 289-293.

The symbols (ρ^i, π_i) will be used for unit contravariant and covariant vectors, respectively, and (r^i, p_i) will be used for general contravariant and covariant vectors.

LEMMA. *The equations*

$$F_{r^i}(x, \rho) - \kappa\pi_i = 0, \quad F(x, \rho) = 1, \quad (\kappa > 0) \qquad (i = 1, \cdots, n), \tag{5.2}$$

in which (π) *is a unit covariant vector at* (x), *can be solved for* κ *and the* ρ's, *thus:*

$$\begin{aligned} \kappa &= \kappa\,(x^1, \cdots, x^n; \pi_1, \cdots, \pi_n) \\ \rho^i &= \rho^i(x^1, \cdots, x^n; \pi_1, \cdots, \pi_n) \end{aligned} \tag{5.3}$$

where the functions κ *and* ρ^i *are of class* C^1 *and the correspondence* (5.3) *is one-to-one between the totality of unit contravariant and covariant vectors* (ρ) *and* (π) *at* (x).

For a proof of essentially this lemma, see Morse, *op. cit.*, p. 241.

(A) *Let functions* $r^i(x, p)$ *be defined by the identities*

$$r^i(x, \tau\pi) = \tau\rho^i(x, \pi) \qquad (\tau > 0), \tag{5.4}$$

(π) *being a unit covariant vector. Then*

$$r^i = r^i(x, p) \tag{5.5}$$

is a one-to-one correspondence between all the covariant and all the contravariant vectors, (p) *and* (r) *respectively, at* (x). *This correspondence agrees with* (5.3) *when* (p) *is a unit vector.*

Now consider the correspondence (5.5) for points $(x) = [\phi(\alpha)]$ on M and for vectors (p) which satisfy the linear homogeneous equations

$$p_i\phi_h{}^i(\alpha) = 0 \qquad (h = 1, \cdots, m). \tag{5.6}$$

(B) We restrict the vectors $\pi^{(j)}(q)$ used in the definition of unit covariant vectors [cf. (5.1)] by the requirement that, when q is at (α) on M, the first $(n - m)$ of these vectors constitute a set $\pi^{(j)}(\alpha)$ $(j = 1, \cdots, n - m)$ of independent solutions of (5.6), so selected as to be of class C^1 in (α).

The unit covariant vector solutions of (5.6) for fixed (α) are

$$\pi_i(\alpha) = \lambda_j\pi_i{}^{(j)}(\alpha), \qquad \lambda_j\lambda_j = 1, \qquad (j = 1, \cdots, n - m). \tag{5.7}$$

This set of vectors is topologically equivalent to an $(n - m - 1)$-sphere. The general solution of (5.6) is $p_i(\alpha) = \tau\lambda_j\pi_i^{(j)}(\alpha)$ as τ ranges over all positive numbers. From the above lemma, together with equations (4.3), (5.6), and (5.7), we see that the transversal contravariant vectors (r) at (α) are the vectors

$$r^i = r^i[\phi(\alpha), \tau\lambda_j\pi^{(j)}(\alpha)], \qquad \lambda_j\lambda_j = 1 \qquad (\tau > 0), \tag{5.8}$$

and our theorem follows at once.

We will say that an extremal [cf. equations (3.12)]

$$x^i = g^i(x_0, \rho_0 s) \tag{5.9}$$

is *cut transversally by M at its initial point,* or *issues transversally from M,* if (x_0) is on M and (ρ_0) is a unit transversal vector at (α). The set of all extremals issuing transversally from M is given, near q_0, by the equations

$$\begin{aligned} x^i &= g^i[\phi(\alpha), sr(\phi, \lambda_j\pi^{(j)}(\alpha))] \\ &= g^i[\phi, r(\phi, s\lambda_j\pi^{(j)})], \end{aligned} \tag{5.10}$$

where

$$\lambda_j\lambda_j = 1, \qquad 0 \leq s < s_0 \tag{5.11}$$

s_0 being a sufficiently small positive quantity.

(C) *The quantities* $(\alpha, \lambda, s) \equiv (\alpha^1, \cdots, \alpha^m, \lambda_1, \cdots, \lambda_{n-m}, s)$ *are a coördinate system near* q_0. *For a fixed set of values* (α), *the quantities* (λ, s) *constitute a polar coördinate system on the* $(n-m)$*-dimensional cone whose elements are the extremals issuing transversally from* M *at* (α). *A set of values* (α, λ) *specifies one of these extremals, on which* s *is the arc length. In terms of* (α, λ, s), *the equation of* M *near* q_0 *is* $s = 0$. *The transformation* (5.10) *between* (x) *and* (α, λ, s) *is of class* C^1 *save when* $s = 0$.

6. Flat transversal cones. A point q on M is generally a conical point of the $(n-m)$-manifold covered by the extremals issuing transversally from M at q. In certain important special cases, which we now investigate, this $(n-m)$-manifold has no conical point at q but has there a unique tangent $(n-m)$-plane. This is equivalent to saying that the transversal vectors at q lie in an $(n-m)$-plane.

THEOREM. *Given m, the transversal vectors to every differentiable m-manifold at any point where the conditions of* § 2 *are fulfilled will always lie in an* $(n-m)$*-plane if and only if the following condition is satisfied:* (1) *in case* $m = n-1$, *that*

$$F(x, r) \equiv F(x, -r), \tag{6.1}$$

(2) *in case* $0 < m < n - 1$, *that* F *be of the form*

$$F(x, r) = \sqrt{a_{ij}(x) r^i r^j}. \tag{6.2}$$

With the aid of the geometric interpretation of transversality [§ 4, (A) and (B)], we see that this theorem is a consequence of the following lemmas applied to the indicatrix, S_{n-1}. We prove the first lemma in a somewhat more general form than necessary, because of its geometric interest.

LEMMA 1. *Let* S_{n-1} *be a differentiable* $(n - m)$*-manifold in an affine* n*-space. Suppose every ray from a certain point,* O*, meets* S_{n-1} *in a single point. Suppose further that, for a given positive integer* $m < n$ *and for every* m*-plane* τ_m *through* O*, the point set* $S_{n-m-1}(\tau_m)$ [For this notation, see § 4 (B)] *is the intersection of* S_{n-1} *with an* $(n - m)$*-plane* $E_{n-m}(\tau_m)$ *through* O*. Then* S_{n-1} *is symmetric in* O *and, if* $m < n - 1$*, is a hyperellipsoid.*

LEMMA 2. *In the same* n*-space, if* S_{n-1} *is a closed convex differentiable* $(n - 1)$*-manifold, symmetric in* O*, then* $S_0(\tau_{n-1})$ *is, for every* τ_{n-1}*, the intersection of* S_{n-1} *with a line through* O*. If* S_{n-1} *is a hyperellipsoid, center at* O*, then* $S_{n-m-1}(\tau_m)$ *is always* $(m = 1, \cdots, n - 1)$*, the intersection of* S_{n-1} *with an* $(n - m)$*-plane through* O*.*

Proof. We prove only Lemma 1, since Lemma 2 presents no difficulty. We give the proof with the aid of six subsidiary results, of which (I)-(IV) are in the present section.

(I) If the hypothesis of Lemma I holds for a given value $m < n - 1$, then it holds with $(m + 1)$ in place of m.

For, let τ_{m+1} be any $(m + 1)$-plane, and let $(\tau_m{}^0, \tau_m{}^1)$ be two non-parallel m-planes on τ_{m+1}. Then $S_{n-m-2}(\tau_{m+1})$ is the common part of $S_{n-m-1}(\tau_m{}^0)$ and $S_{n-m-1}(\tau_m{}^1)$ and is therefore the intersection of S_{n-1} with the linear space E in which $E_{n-m}(\tau_m{}^0)$ and $E_{n-m}(\tau_m{}^1)$ intersect. It is easy to verify that $S_{n-m-2}(\tau_{m+1})$, as the set of points where the tangent $(n - 1)$-plane to S_{n-1} contains a parallel to τ_{m+1}, is of dimensionality at least $(n - m - 2)$. Hence E is of dimensionality at least $(n - m - 1)$. We must also show that E is of dimensionality at most $(n - m - 1)$, in other words that $E_{n-m}(\tau_m{}^0)$ and $E_{n-m}(\tau_m{}^1)$ do not coincide. This follows from the fact that, if E_{n-1} is an $(n - 1)$-plane containing $\tau_m{}^0$ and not containing $\tau_m{}^1$, the points where the tangent plane to S_{n-1} is parallel to E_{n-1} belong to $S_{n-m-1}(\tau_m{}^0)$ and not to $S_{n-m-1}(\tau_m{}^0)$.

(II) As a consequence of result (I), it is sufficient to prove Lemma 1 in the two cases $m = n - 1, n - 2$.

The following statements are easy to verify.

(III) Under the hypothesis of Lemma 1, for any $(n-m)$-plane, E_{n-m}, through O, there exists an m-plane, τ_m, through O such that

$$E_{n-m} = E_{n-m}(\tau_m). \tag{6.3}$$

(IV) Let S^*_1 be the curve in which S_{n-1} is met by an arbitrary given 2-plane, E_2, through O. Lemma 1 will follow in all its generality if we show that S^*_1 is symmetric in O and, in case $m = n-2$, is an ellipse.

We first note that no $(n-1)$-plane, τ_{n-1}, through O can be tangent to S_{n-1}. For, by the hypothesis of Lemma 1 in the case $m = n-1$, $S_0(\tau_{n-1})$ consists of exactly two points. The tangent planes at these points can be obtained by moving an $(n-1)$-plane, keeping it parallel to τ_{n-1}, in either direction from τ_{n-1} to the last positions in which it meets S_{n-1}. Hence neither point of $S_0(\tau_{n-1})$ lies on τ_{n-1}. It follows that no ray from O can be tangent to S_{n-1}. For if Q were a point of tangency of such a ray, then the $(n-1)$-plane tangent to S_{n-1} at Q would contain the ray and hence pass through O. We can accordingly represent the curve S^*_1 in terms of polar coördinates (ρ, ϕ) on E_2 with the pole at 0, by an equation

$$\rho = K(\phi), \tag{6.4}$$

where $K(\phi)$ is differentiable. Furthermore,

$$K(\phi + 2\pi) = K(\phi). \tag{6.5}$$

From the hypotheses of the lemma it follows that the tangents to S^*_1 at the opposite ends of any chord through O are parallel. This condition can be expressed in the form

$$\frac{d}{d\phi} \log K(\phi) = \frac{d}{d\phi} \log K(\phi + \pi). \tag{6.6}$$

From equations (6.5) and (6.6), it follows that $K(\phi)$ is periodic of period π, hence that S^*_1 is symmetric in O.

Lemma 1 *is now established in the case* $m = n-1$.

7. The case of an ellipsoidal indicatrix.[6] The present section completes the work of § 6 by establishing § 6, Lemma 1, in the case $m = n-2$ [cf. § 6 (II)]. In this case, (6.3) becomes

---

[6] The work of this section is largely a generalization of a proof in the case $(m, n) = (1, 3)$ by W. Blaschke, "Räumliche Variationsprobleme mit symmetrischen Transversalitätsbedingungen," *Leipzig. Berichte*, vol. 68 (1916), pp. 50-55.

(7.1) $$E_2 = E_2(\tau_{n-2})$$

and, by our definitions,

(7.2) $$S^*_1 = S_1(\tau_{n-2}).$$

Let τ^0_{n-1} be an $(n-1)$-plane containing τ_{n-2}. By § 6 (I) and the lemma for the case $m = n-1$, the points where the tangent plane to S_{n-1} is parallel to τ^0_{n-1} are the end points (P, Q) of a chord of S_{n-1} through O.

(V) Let τ_{n-1} be the $(n-1)$-plane parallel to τ^0_{n-1} through a variable point, O', of the chord PQ. Then the intersection, S_{n-2}, of S_{n-1} with τ_{n-1} satisfies, in the space τ_{n-1}, the hypothesis of § 6, Lemma 1, with $(n-2, n-1)$ in place of (m, n) and with O' in place of O.

Proof. We first regard τ_{n-2} as free to assume any position through O on the fixed $(n-1)$-plane τ^0_{n-1}. Then $E_2(\tau_{n-2})$ always passes through (P, Q). Conversely, if $E_2 = E_2(\tau_{n-2})$ is any 2-plane through (P, Q), then $\tau_{n-2} \subset \tau^0_{n-1}$. As a consequence of our definitions, the intersection of $S_1(\tau_{n-2})$ with S_{n-2} is the set of points where the tangent $(n-2)$-plane to S_{n-2} is parallel to τ_{n-2}. These are also the points in which S_{n-2} is met by the line common to τ_{n-1} and $E_2(\tau_{n-2})$. For the complete establishment of result (V), it remains only to show that this line, as τ_{n-1} varies, always meets $S_1(\tau_{n-2})$ in just two points (A, B). When τ_{n-1} is at τ^0_{n-1}, this follows from the definition of S_{n-1}. As τ_{n-1} moves from τ^0_{n-1}, towards P for example, (A, B) may be thought of as varying continuously. If there were a first position in which the line AB met $S_1(\tau_{n-2})$ in a third point, C, we should then have AB tangent to $S_1(\tau_{n-2})$ at C. But C would then be a point where the tangent to S_{n-1} is parallel to τ_{n-1} and, by § 6 (I) and Lemma 1 for $n-m=1$, (P, Q) are the only such points. The proof of result (V) is now complete

(VI) In the case $m = n-2$, $S^*_1 = S_1(\tau_{n-2})$ is an ellipse.

Proof. From the proof of (V), together with § 6, Lemma 1, in the case $n-m=1$, we see that each chord of the set AB of parallel chords of $S_1(\tau_{n-2})$ is bisected by PQ. We refer to PQ as a *line of symmetry* and to the common direction of the chords AB as the *corresponding direction.* Now, keeping τ_{n-2} fixed, let τ^0_{n-1} adopt every position such that $\tau^0_{n-1} \supset \tau_{n-2}$. Then PQ varies correspondingly and adopts the position of every line in E_2 through O. Accordingly, every such line is a line of symmetry of $S_1(\tau_{n-2})$. Let (u, v) be a coördinate system in E_2, origin at O, such that the u-axis is in the corresponding direction to the v-axis, regarded as a line of symmetry. Let (L, L') be a second pair of lines through O, where L' is in the corresponding direction to L and where (L, L') are of positive and negative slopes respectively with reference to the (u, v)-system. By a compression in the v-direction, we can transfer to a new coördinate system (u, v'), in terms of which L and L' have negative reciprocal slopes. Let ζ be the inclination of L, figured as if (u, v') were a rectangular cartesian coördinate system, and suppose L so chosen that ζ is an irrational multiple of π. Let (ρ, ϕ) be the polar coördinate system superposed in the usual way on (u, v'). Then, using $\rho = K(\phi)$ to represent $S_1(\tau_{n-2})$, we have, by symmetry in the lines $\phi = \pi/2$ and $\phi = \zeta$,

$$K(\phi) = K(\pi - \phi), \qquad K(\phi) = K(2\zeta - \phi). \tag{7.3}$$

Hence

$$K(\phi) = K(\pi - 2\zeta + \phi). \tag{7.4}$$

But $K(\phi)$ is also periodic of period π, and the periods π and $\pi - 2\zeta$ are not rational multiples of one another. Hence K is a constant and $S_1(\tau_{n-2})$ is a circle about O in terms of the (u, v')-system. Therefore, in our affine n-space, $S^*_1 = S_1(\tau_{n-2})$ is an ellipse. By § 6 (IV), this completes the proof.

8. Normal coördinates with respect to M. (A) *If* $F(x, r) \equiv F(x, -r)$, *then, in the notation of* (5. 5),

$$r^i(x, p) \equiv r^i(x, -p). \tag{8.1}$$

If $F(x, r) = \sqrt{a_{ij}(x) r^i r^j}$, *then* $r^i(x, p)$ *is linear homogeneous in* $(\bar\kappa p_1, \cdots, \bar\kappa p_n)$, *where* $\bar\kappa$ *is the function* κ *of equations* (5. 3) *figured for the unit covariant vector at* (x) *in the direction of* (p).

Result (A) can be directly verified in equations (5. 2)–(5. 4).

Our *normal coördinates* (z) *with respect to* M *near* q_0 are defined by the transformation

$$x^i = G^i(z), \tag{8.2}$$

where $G^i(z)$ is obtained by substituting $(z^1, \cdots, z^m)$ for $(\alpha^1, \cdots, \alpha^m)$ and $(z^{m+1}, \cdots, z^n)$ for $(s\lambda_1, \cdots, s\lambda_{n-m})$, respectively, in the functions g^i of equations (5. 10).

(B) *In general, the functions* $G^i(z)$ *are of class*[7] C^1 *save when* $z^{m+1} = \cdots = z^n = 0$. *If the condition of* § 6, *Theorem, is fulfilled, these functions are of class* C^1 *without exception and have a non-zero jacobian throughout some neighborhood of* q_0.

The first part of (B) follows from the work of § 5 and the second part follows from (A) above.

---

[7] Or of higher class, if we strengthen the differentiability assumptions on (M, R, F).

In terms of the normal coördinates (z), M is defined near q_0 by the equations

$$z^{m+1} = \cdots = z^n = 0, \tag{8.3}$$

and the extremals issuing transversally from a point $(z) = (a^1, \cdots, a^m, 0, \cdots, 0)$ of M near q_0 are defined, near M, by the equations

$$\begin{aligned} &z^i = a^i & & (i = 1, \cdots, m), \\ &z^{m+j} = \lambda_j s \quad (0 \leq s < s_0) & & (j = 1, \cdots, n-m), \end{aligned} \tag{8.4}$$

where s represents arc length and where

$$\lambda_j \lambda_j = 1. \tag{8.5}$$

If the condition of § 6, Theorem, is satisfied, we have the reversible case, and all the extremals which are cut transversally by M, not necessarily at their initial points, are given near q_0 by equations (8.4), with s permitted to take on negative as well as positive values. If the condition of § 6, Theorem, is not satisfied, then in general an extremal cut transversally by M at a point q will be of class C^2 in terms of (z) save for a corner at q, and one of the two parts into which q divides it will be a ray in (z)-space.

QUEENS COLLEGE,
FLUSHING, N. Y.

PROBLEMS OF CLOSEST APPROXIMATION ON A TWO-DIMENSIONAL REGION.*

By Dunham Jackson.

---

1. Introduction.

The theorems of Markoff and Bernstein on the derivatives of polynomials and trigonometric sums can be made to serve as basis for a theory of the convergence of certain types of polynomial and trigonometric approximation to functions of a single real variable.[1] It is fairly apparent that similar methods can be applied to functions of two variables. In the carrying out of this process there are so many possible variations and combinations that it would not be profitable to enumerate the resulting theorems systematically, still less to discuss them successively in detail. Nevertheless, the extension involves some adjustments which are not entirely automatic or superficial, and it has been thought worth while to present below some typical results illustrating the differences between the one-dimensional and two-dimensional formulations.[2]

There is occasion first to develop two-dimensional versions of the Markoff and Bernstein theorems themselves.

2. Theorems of Markoff and Bernstein for polynomials. Let $P(x, y)$ be a polynomial of the n-th degree[3] in the variables x and y together. (It is then of the n-th degree in each variable separately; on the other hand, an arbitrary polynomial of degree n in each variable is of degree $2n$ in both together, and the discussion as given is applicable on replacement of n by $2n$.)

If $| P(x, y) | \leqq L$ at all points of a straight line segment of length h in the (x, y)-plane, and if $\partial P/\partial s$ denotes directional differentiation along the line,

$$\left| \frac{\partial P}{\partial s} \right| \leqq \frac{2n^2 L}{h}$$

---

* Received November 23, 1937.

[1] See for example D. Jackson, "Certain problems of closest approximation," *Bulletin of the American Mathematical Society*, vol. 39 (1933), pp. 889-906; "Bernstein's theorem and trigonometric approximation," *Transactions of the American Mathematical Society*, vol. 40 (1936), pp. 225-251.

[2] See also E. Carlson, "On the convergence of trigonometric approximations for a function of two variables," *Bulletin of the American Mathematical Society*, vol. 32 (1926), pp. 639-641; E. L. Mickelson, "On the approximate representation of a function of two variables," *Transactions of the American Mathematical Society*, vol. 33 (1931), pp. 759-781.

[3] This expression will be understood throughout to mean "of the n-th degree at most."

on the entire segment, and

$$\left| \frac{\partial P}{\partial s} \right| \leqq \frac{nL}{[\delta(h-\delta)]^{\frac{1}{2}}}$$

at a point of the segment whose distance from the nearer end is δ. For a transformation of coördinates by translation and rotation to a new system with one axis along the specified line reduces $P(x, y)$ on the line to a polynomial of the n-th degree in a single variable, having L as an upper bound for its absolute value on the segment, and having $\partial P/\partial s$ for its derivative with respect to the variable, so that the upper bounds for $|\, \partial P/\partial s \,|$ are given by the standard forms of the Markoff and Bernstein theorems respectively.

If $|\, P(x, y) \,| \leqq L$ throughout the square $-1 \leqq x \leqq 1$, $-1 \leqq y \leqq 1$, application of Markoff's theorem to the polynomials in one variable obtained by holding the other variable fast gives $|\, \partial P/\partial x \,| \leqq n^2 L$, $|\, \partial P/\partial y \,| \leqq n^2 L$ on the square. If $\partial P/\partial s$ is the directional derivative at any point of the square in a direction making an arbitrary angle α with the x-axis,

$$\left| \frac{\partial P}{\partial s} \right| = \left| \frac{\partial P}{\partial x} \cos \alpha + \frac{\partial P}{\partial y} \sin \alpha \right| \leqq n^2 L(|\cos \alpha| + |\sin \alpha|) \leqq 2^{\frac{1}{2}} n^2 L.$$

(Here and in the next statement it is sufficient, as regards the degree of $P(x, y)$, that it be of the n-th degree in each variable separately.) At a point of the square whose shortest distance from the boundary is δ, Bernstein's theorem similarly gives[4]

$$\left| \frac{\partial P}{\partial x} \right| \leqq \frac{nL}{[\delta(2-\delta)]^{\frac{1}{2}}}, \quad \left| \frac{\partial P}{\partial y} \right| \leqq \frac{nL}{[\delta(2-\delta)]^{\frac{1}{2}}}, \quad \left| \frac{\partial P}{\partial s} \right| \leqq \frac{2^{\frac{1}{2}} nL}{[\delta(2-\delta)]^{\frac{1}{2}}}.$$

It is perhaps not necessary at the present stage to formulate the corresponding statements for a rectangle of arbitrary size, shape, and orientation.

Suppose that a point D lies on two mutually perpendicular line segments, on each of which $|\, P(x, y) \,| \leqq L$, the length of each segment being $\geqq h$. If ξ, η are coördinates with respect to a pair of axes along the two lines,

$$|\, \partial P/\partial \xi \,| \leqq 2n^2 L/h, \qquad |\, \partial P/\partial \eta \,| \leqq 2n^2 L/h$$

at the point D, and if $\partial P/\partial s$ is any directional derivative at the point,

$$|\, \partial P/\partial s \,| \leqq 2^{3/2} n^2 L/h.$$

---

[4] In case the horizontal and vertical distances from the boundary are different it is to be noted that if $0 < \delta < \delta' \leqq 1$,

$$\delta(2-\delta) < \delta'(2-\delta'),$$

the expression $\delta(2-\delta)$ as a function of δ having its maximum for $\delta = 1$.

More generally, suppose that the segments are oblique to each other, and let γ denote the magnitude of the acute angle formed by their lines (even if the segments themselves terminate at D and form an obtuse angle there). Let ξ, η be coördinates referred to a pair of rectangular axes with origin at D, the ξ-axis extending along one of the given lines, let P_γ denote the result of differentiation along the other given line, and let P_s be an arbitrary directional derivative at D, in a direction making an angle α with the ξ-axis. Then, with an appropriate choice of the sense of differentiation on each line (the discussion being concerned essentially only with the *magnitudes* of the derivatives),

$$\begin{gathered} |P_\xi| \leqq 2n^2L/h, \qquad |P_\gamma| \leqq 2n^2L/h, \\ P_\gamma = P_\xi \cos\gamma + P_\eta \sin\gamma, \qquad P_\eta = [-P_\xi \cos\gamma + P_\gamma]/\sin\gamma, \\ P_s = P_\xi \cos\alpha + P_\eta \sin\alpha = [P_\xi \sin(\gamma - \alpha) + P_\gamma \sin\alpha]/\sin\gamma, \\ |P_s| \leqq 4n^2L/(h\sin\gamma). \end{gathered}$$

Let R be a closed region (more generally, any point set) for which there are two positive constants h, γ such that each point of R lies on two line segments belonging entirely to R, each of length $\geqq h$, their lines making with each other a minimum angle $\geqq \gamma$. If $|P(x, y)| \leqq L$ throughout R,

$$|\partial P/\partial s| \leqq 4n^2L/(h\sin\gamma)$$

at all points of R (including the boundary),[5] for all directions of differentiation.

If $|P(x, y)| \leqq L$ on two segments which cross at D, the length of each being $\geqq h$, and the angle between them γ, as before, and if D is at a distance $\geqq \delta$ from each end of each segment,

$$(1) \qquad \left|\frac{\partial P}{\partial s}\right| \leqq \frac{2nL}{[\delta(h - \delta)]^{\frac{1}{2}} \sin\gamma}$$

for differentiation in any direction at D. If $|P(x, y)| \leqq L$ throughout a region R of the sort described in the last paragraph, and if $0 < \delta \leqq \frac{1}{2}h$, (1) holds for an arbitrary directional derivative at any point of R whose minimum distance from the boundary is $\geqq \delta$. Somewhat less explicitly, $|\partial P/\partial s| \leqq CnL/\delta^{\frac{1}{2}}$, where C depends only on h and γ. If R_1 is a closed point set interior to R, there is a constant C' for R and R_1 such that $|\partial P/\partial s| \leqq C'nL$ throughout R_1.

3. Theorems of Markoff and Bernstein for trigonometric sums. In the preceding discussion repeated use has been made of the fact that a transformation of coördinates carries a polynomial of the n-th degree into a polynomial of the n-th degree. This property is not shared by trigonometric sums in two variables, and for application to such functions the argument has to be reconsidered accordingly.

---

[5] Both hypothesis and conclusion with regard to $P(x, y)$ are such that if they hold at interior points of R they necessarily hold on the boundary.

Some conclusions, to be sure, which involve no rotation of axes, can be obtained immediately from the corresponding theorems in one variable. Let $T(x, y)$ be a trigonometric sum in x and y which is of the n-th order (i. e., of the n-th order at most) in each variable separately. If $| T(x, y) | \leqq L$ for all values of x and y, $| \partial T/\partial x | \leqq nL$, $| \partial T/\partial y | \leqq nL$, and $| \partial T/\partial s | \leqq 2^{\frac{1}{2}}nL$ for differentiation in any direction at any point. If $| T(x, y) | \leqq L$ throughout a rectangle with sides parallel to the coördinate axes, $| \partial T/\partial s | \leqq Cn^2L$ throughout the rectangle, the constant C depending only on the dimensions of the rectangle,[6] and $| \partial T/\partial s |$ has a constant multiple of nL as upper bound in any closed region interior to the rectangle.[7]

Suppose now that $T(x, y)$ is of the n-th order in the two variables together, and that $| T(x, y) | \leqq L$ on a line segment of length h *with rational slope* p/q, the numbers p and q being integers and relatively prime. Let (x_0, y_0) be a point of the segment, let $(p^2 + q^2)^{\frac{1}{2}} = r$, and let

$$\xi = (q/r^2)(x - x_0) + (p/r^2)(y - y_0),$$
$$\eta = (p/r^2)(x - x_0) - (q/r^2)(y - y_0).$$

The given segment lies in the line $\eta = 0$. The inverse transformation is

$$x = x_0 + q\xi + p\eta,$$
$$y = y_0 + p\xi - q\eta.$$

A trigonometric sum of order n in x and y is for $\eta = 0$ a trigonometric sum of order np_0 in ξ, if p_0 is the larger of p and q; a term $\cos n_1x \cos n_2y$, for example, where $n_1 + n_2 = n$, becomes

$$\cos n_1(x_0 + q\xi) \cos n_2(y_0 + p\xi)$$

of order $n_1q + n_2p \leqq np_0$. If Δs denotes distance in the (x, y) plane, $\Delta s = (\Delta x^2 + \Delta y^2)^{\frac{1}{2}} = r\Delta\xi$ on the line $\eta = 0$, and the difference between the values of ξ corresponding to the ends of the given segment is h/r. Consequently[8]

$$\left| \frac{\partial T}{\partial \xi} \right| \leqq G_0n^2L, \qquad \left| \frac{\partial T}{\partial s} \right| = \left| \frac{1}{r} \frac{\partial T}{\partial \xi} \right| \leqq Gn^2L,$$

---

[6] See for example D. Jackson, *Transactions, loc. cit.*, p. 230, Theorem 2.

[7] *Ibid.*, p. 227, Theorem 1.

[8] *Transactions, loc. cit.*, Theorem 2.

when G_0 and G depend only on h and r (since $p_0 \leqq r$). At a point of the segment whose distance from the nearer end as measured in the (x, y) plane is not less[9] than δ,

$$\left| \frac{\partial T}{\partial s} \right| \leqq \frac{G_1 n L}{[\delta(h-\delta)]^{\frac{1}{2}}},$$

the constant G_1 also depending only on h and r. As $h - \delta \geqq \frac{1}{2}h$ it may be omitted from the denominator, with a corresponding change in the value of G_1.

The derivative in an arbitrary direction at a point common to two segments making an angle with each other, on each of which $| T(x, y) | \leqq L$, can then be dealt with as in the polynomial case.

Let R' be a region (or more general point set) for which there are three constants $h > 0$, $\gamma' > 0$, r_0 such that each point of R' lies on two line segments belonging to R', each of length $\geqq h$ and with a rational slope p/q satisfying the condition that $p^2 + q^2 \leqq r_0^2$, while the lines of the segments make with each other a minimum angle $\geqq \gamma'$. If $| T(x, y) | \leqq L$ throughout R', $| \partial T/\partial s | \leqq C n^2 L$ for differentiation in an arbitrary direction at any point of R', the constant C depending only on h, r_0, and γ'; since the hypothesis admits only a finite number of pairs of values of p and q, and so a finite number of values of $(p^2 + q^2)^{\frac{1}{2}} = r$, the various constants G of the second paragraph preceding which correspond to a fixed h and different values of r can be replaced by a single one equal to the largest of them. At a point whose minimum distance from the boundary is $\geqq \delta$, $0 < \delta \leqq \frac{1}{2}h$,

$$| \partial T/\partial s | \leqq C_1 n L/\delta^{\frac{1}{2}},$$

with a value of C_1 depending only on h, r_0, and γ'. As in the second paragraph preceding the manner of dependence of the constants G and G_1 on h was left wholly unspecified, it is to be noted that under the present hypotheses any segment entering into the discussion whose length is greater than h can be replaced by a segment of length h exactly. If R_1 is a closed point set interior to R', there is a constant C' determinate for R' and R_1 such that $| \partial T/\partial s | \leqq C' n L$ throughout R_1.

The conditions imposed on R in the preceding section and those on R' here will be satisfied by *any region R for which there are two positive constants h, γ such that every point of R is vertex of a circular sector of radius $\geqq h$ and angle $\geqq \gamma$, belonging entirely to R.* A value of r_0 meeting the requirements in the case of R' can be associated with any value of $\gamma' < \gamma$.

---

[9] *Ibid.*, Theorem 1.

Some of the results obtained above will be used in the following section, while others have been inserted here merely for purposes of comparison.

4. Approximation by trigonometric sums and polynomials. Trigonometric approximation over the entire plane for a function periodic in both variables has been treated from the point of view of the present article by another writer[10] The ordinary form of Bernstein's theorem in a single variable, applied with respect to x and y separately, can be used to justify the following assertion:

If $f(x, y)$ is a continuous function which is of period 2π in each variable, if $T_n(x, y)$ is a trigonometric sum of the n-th order in each variable (not necessarily of the n-th order in both together), if

$$G_{ns} = \int_{-\pi}^{\pi}\int_{-\pi}^{\pi} | f(x, y) - T_n(x, y) |^s \, dxdy,$$

s being any positive exponent, and if there exists a trigonometric sum $t_n(x, y)$, of the n-th order in each variable, such that

$$| f(x, y) - t_n(x, y) | \leqq \epsilon_n$$

everywhere, then

$$| f(x, y) - T_n(x, y) | \leqq 4(4n^2 G_{ns})^{1/s} + 5\epsilon_n$$

for all values of x and y.

The proof is closely parallel to that of a corresponding proposition in one variable,[11] the most notable difference, giving rise to the factor $4n^2$ inside the parentheses in the last inequality, being that an interval $| x - x_0 | \leqq 1/(2n)$, of length $1/n$, which enters into the proof in the case of one variable, is to be replaced here by a square $| x - x_0 | \leqq 1/(4n)$, $| y - y_0 | \leqq 1/(4n)$, of area $1/(4n^2)$. (More generally, if the trigonometric sums are of order m in one variable and of order n in the other, n^2 is to be replaced by mn.) If $T_n(x, y)$ is chosen among all trigonometric sums of the order indicated so as to minimize the integral G_{ns}, the fact that G_{ns} does not exceed the corresponding integral with $T_n(x, y)$ replaced by $t_n(x, y)$ implies that $G_{ns} \leqq 4\pi^2\epsilon_n^s$, and $| f(x, y) - T_n(x, y) |$ consequently does not exceed a constant multiple of $n^{2/s}\epsilon_n$. This in conjunction with theorems establishing the existence of trigonometric sums $t_n(x, y)$ giving a specified degree of approximation[12] leads to conclusions as to the uniform convergence of the minimizing sum $T_n(x, y)$

---

[10] See E. Carlson, *loc. cit.*

[11] See D. Jackson, *Bulletin*, *loc. cit.*, pp. 899-900, Lemma 5.

[12] See E. L. Mickelson, *loc. cit.*, § 4.

toward $f(x, y)$. The conclusions can be generalized by the introduction of a weight function and by the use of Hölder's inequality in connection with the integral to be minimized.[13] Similar remarks apply to other propositions which are to be formulated below.

If the property of periodicity is dropped, and the hypotheses are made to refer to a rectangle $a \leqq x \leqq b$, $c \leqq y \leqq d$, the trigonometric sums $T_n(x, y)$ and $t_n(x, y)$ being replaced by polynomials $P_n(x, y)$ and $p_n(x, y)$, of the n-th degree in each variable, with

$$G_{ns} = \int_c^d \int_a^b | f(x, y) - P_n(x, y) |^s \, dx dy,$$

it is found that

$$| f(x, y) - P_n(x, y) | \leqq 4 \left[\frac{64 n^4 G_{ns}}{(b-a)(d-c)} \right]^{1/s} + 5\epsilon_n$$

throughout the rectangle. The factor n^4 in the right-hand member of the inequality, in place of n^2, results from the use of Markoff's theorem instead of that of Bernstein.

Let the domain of integration more generally be an arbitrary closed region R possessing the sector property described at the end of the preceding section. Let $f(x, y)$ be a function continuous throughout R, let $P_n(x, y)$ be a polynomial of the n-th degree in the two variables together, let

$$G_{ns} = \iint_R | f(x, y) - P_n(x, y) |^s \, dx dy,$$

s being a positive exponent, as before, and let it be supposed that there is a polynomial $p_n(x, y)$, of the n-th degree in the two variables together, such that

$$| f(x, y) - p_n(x, y) | \leqq \epsilon_n$$

at all points of R.

Under these conditions the corresponding form of Markoff's theorem obtained in § 2 is applicable. To give the proof in this single instance in some detail, let

$$f(x, y) - p_n(x, y) = r_n(x, y), \qquad | r_n(x, y) | \leqq \epsilon_n,$$
$$P_n(x, y) - p_n(x, y) = \pi_n(x, y),$$

so that

$$f(x, y) - P_n(x, y) = r_n(x, y) - \pi_n(x, y).$$

Let μ_n be the maximum of $| \pi_n(x, y) |$ in R, taken on at a point (x_0, y_0). At any point (x, y) which belongs to the sector associated with (x_0, y_0) by the

---

[13] See e. g. D. Jackson, *Bulletin, loc. cit.*, pp. 901-902, Theorems 9, 10.

hypothesis, and is distant from (x_0, y_0) by not more than $h_n = h\sin\gamma/(8n^2)$, since $|\partial\pi_n/\partial s| \leqq 4n^2\mu_n/(h\sin\gamma)$ throughout R and the entire segment joining (x, y) with (x_0, y_0) belongs to R,

$$|\pi_n(x, y) - \pi_n(x_0, y_0)| \leqq \mu_n/2, \qquad |\pi_n(x, y)| \geqq \mu_n/2.$$

If $\mu_n \geqq 4\epsilon_n$, which means that $|r_n(x, y)| \leqq \mu_n/4$,

$$|r_n(x, y) - \pi_n(x, y)| \geqq \mu_n/4$$

at the same points. As this relation holds at least throughout a sector of radius h_n, angle γ, and area $\gamma h_n^2/2$, belonging to R,

$$G_{ns} \geqq \frac{\gamma h_n^2}{2}\left(\frac{\mu_n}{4}\right)^s = \frac{h^2\gamma\sin^2\gamma}{128n^4}\left(\frac{\mu_n}{4}\right)^s, \qquad \mu_n \leqq 4\left(\frac{128n^4 G_{ns}}{h^2\gamma\sin^2\gamma}\right)^{1/s}.$$

Otherwise $\mu_n < 4\epsilon_n$. As μ_n certainly can not exceed the sum of the two alternative upper bounds, while $|\pi_n(x, y)| \leqq \mu_n$ and $|r_n(x, y)| \leqq \epsilon_n$, it is true in either case that

$$(2) \qquad |f(x, y) - P_n(x, y)| = |r_n(x, y) - \pi_n(x, y)| \leqq 4\left(\frac{128n^4 G_{ns}}{h^2\gamma\sin^2\gamma}\right)^{1/s} + 5\epsilon_n$$

throughout R.

Theorems on the existence of approximating polynomials $p_n(x, y)$ with specified upper bounds of error, to be sure, are for approximation over a rectangular region.[14] For direct application of those theorems it is to be assumed that $f(x, y)$ satisfies the requisite conditions of continuity or differentiability on a rectangle containing R, or at least that its definition in R can be so extended that the conditions hold throughout such a rectangle.[15]

In the analogous problem of trigonometric approximation on a region R having the sector property the conclusion corresponding to (2) is that

$$|f(x, y) - T_n(x, y)| \leqq C_s(n^4 G_{ns})^{1/s} + 5\epsilon_n,$$

where C_s, though not so easy to calculate explicitly as in the polynomial case, is again dependent only on h, γ, and s.

Let the phrase "mixed sum" be used to describe a function of x and y which is a polynomial with respect to one variable and a trigonometric sum with respect to the other. Let $f(x, y)$ be a continuous function of the two variables for $a \leqq x \leqq b$ and for all values of y, of period 2π with respect to y, let $U_n(x, y)$ be a mixed sum which is a polynomial of the n-th degree in x and a trigonometric sum of the n-th order in y, and let

---

[14] See D. Jackson, *Über die Genauigkeit der Annäherung stetiger Funktionen . . .*, Dissertation, Göttingen, 1911, pp. 88-95; E. L. Mickelson, *loc. cit.*, § 5.

[15] In this connection see Hassler Whitney, "Analytic extensions of differentiable functions defined in closed sets," *Transactions of the American Mathematical Society*, vol. 36 (1934), pp. 63-89.

The problem of minimizing G_{ns} can be regarded alternatively as that of minimizing the integral over a rectangle containing R, with a weight function equal to 1 in R and equal to 0 elsewhere.

$$G_{ns} = \int_{-\pi}^{\pi} \int_{a}^{b} | f(x, y) - U_n(x, y) |^s \, dxdy.$$

Suppose that there exists a mixed sum $u_n(x, y)$, of the n-th degree in x and of the n-th order in y, such that

$$| f(x, y) - u_n(x, y) | \leqq \epsilon_n$$

for $a \leqq x \leqq b$ and for all values of y. By the use of Markoff's theorem (in a single variable) for differentiation with respect to x and of Bernstein's theorem for differentiation with respect to y it is found that

$$| f(x, y) - U_n(x, y) | \leqq 4 \left(\frac{16 n^3 G_{ns}}{b - a} \right)^{1/s} + 5\epsilon_n$$

throughout the strip of the (x, y) plane under consideration, the factor n^3 being associated with G_{ns} here instead of n^2 or n^4. Theorems are available [16] on the existence of approximating sums $u_n(x, y)$ corresponding to specified orders of magnitude of ϵ_n.

A trigonometric substitution can be used in connection with polynomial approximation in the same manner as in the one-dimensional case [17] to improve the order of magnitude obtained for the upper bound of error in the interior of the domain of integration. The substitution can be applied either to one variable or to both. In the hypotheses of an earlier paragraph relating to polynomial approximation over a rectangle, let the rectangle be taken for simplicity as the square $-1 \leqq x \leqq 1$, $-1 \leqq y \leqq 1$. Let N be the smallest integer $\geqq 1/s$. By setting $x = \cos\theta$, $y = \cos\phi$, and using the results obtained above for trigonometric approximation over the entire plane, it may be shown that

$$(3) \qquad | f(x, y) - P_n(x, y) | \leqq \frac{4[16(n + N)^2 G_{ns}]^{1/s} + 5\epsilon_n}{[(1 - x^2)(1 - y^2)]^{N/2}}.$$

The significant difference between this and the earlier result for polynomials is the replacement of n^4 by a factor of the order of n^2, with compensating introduction of a denominator which vanishes on the boundary of the square. If the substitution is made for just one variable, say $y = \cos\phi$, and combined with the conclusion of the last paragraph for approximation by mixed sums, it is found that

---

[16] E. L. Mickelson, *loc. cit.*, § 6.

[17] See e. g. D. Jackson, *Bulletin*, *loc. cit.*, p. 905, Lemma 8.

$$| f(x, y) - P_n(x, y) | \leqq \frac{4[13(n+N)^3 G_{ns}]^{1/s} + 5\epsilon_n}{(1-y^2)^{N/2}}.$$

This inequality, with a factor of the order of n^3, is effective on two sides of the square as well as in the interior. It appears then, as far as the evidence of the present reasoning goes, that the factor n^4 is required at the corners of the square, a factor of the order of n^3 at points of the sides other than the corners, and a factor of the order of n^2 in the interior. Either of the last two inequalities can be adapted to a rectangle of arbitrary dimensions with sides parallel to the axes by linear transformation of the variables separately. Expressed in terms of the degree of the polynomials in the two variables jointly, they can be carried over by rotation of axes to a rectangle of arbitrary orientation.

For a square of side $2h$ with middle point at (x_0, y_0) the inequality (3), evaluated at the middle point, becomes

$$| f(x_0, y_0) - P_n(x_0, y_0) | \leqq 4[16(n+N)^2 G_{ns}/h^2]^{1/s} + 5\epsilon_n.$$

If the hypotheses are stated for a circle of radius δ and center (x_0, y_0), with G_{ns} now representing the value of the integral over the circle,

$$(4) \qquad | f(x_0, y_0) - P_n(x_0, y_0) | \leqq 4[32(n+N)^2 G_{ns}/\delta^2]^{1/s} + 5\epsilon_n,$$

since the circle contains a square of side $2^{\frac{1}{2}}\delta$ with sides parallel to the axes. If the definition of G_{ns} and the other hypotheses are formulated for an arbitrary closed region R, it follows that the right-hand member of (4) is an upper bound for $| f(x, y) - P_n(x, y) |$ at any point of R whose minimum distance from the boundary is $\geqq \delta$. For the dependence of the quantity in brackets on n a factor of the order of n^2 is thus sufficient at any interior point of R. The significance of this observation for the theory of convergence of polynomials of closest approximation is brought out more explicitly by the following inference from it:

If $f(x, y)$ is continuous throughout a closed region R, and if $P_n(x, y)$ is determined among all polynomials of the n-th degree in each variable so as to minimize the integral

$$G_{ns} = \iint\limits_R | f(x, y) - P_n(x, y) |^s \, dxdy,$$

$P_n(x, y)$ will converge toward $f(x, y)$ at every interior point of R, uniformly throughout any closed region interior to R, as n becomes infinite, if there exist polynomials $p_n(x, y)$ of corresponding degree such that

$$| f(x, y) - p_n(x, y) | \leqq \epsilon_n$$

throughout R, with $\lim_{n\to\infty} n^{2/s}\epsilon_n = 0$.

This statement, like others of similar character, can be generalized by introduction of a weight function.

For an arbitrary interior point of an arbitrary rectangle $a \leqq x \leqq b$, $c \leqq y \leqq d$, with sides parallel to the axes, (3) takes the form

$$| f(x,y) - P_n(x,y) | \leqq \frac{C_s[(n^2 G_{ns})^{1/s} + \epsilon_n]}{[(b-x)(x-a)(d-y)(y-c)]^{N/2}}$$

where C_s depends only on s and on the dimensions of the rectangle. By successive applications, after the manner of a corresponding proof in one dimension,[18] this can be made to yield a theorem of similar form with regard to trigonometric approximation. The four overlapping intervals of the argument in one variable are to be replaced here by sixteen overlapping rectangles. It is to be supposed that $f(x,y)$ is continuous throughout a rectangle $a \leqq x \leqq b$, $c \leqq y \leqq d$, where $0 < b - a < 2\pi$, $0 < d - c < 2\pi$, and that there exists a trigonometric sum $t_n(x,y)$ of the n-th order in each variable such that $| f(x,y) - t_n(x,y) | \leqq \epsilon_n$ throughout the rectangle. Then if $T_n(x,y)$ is an arbitrary trigonometric sum of similar order, if s is an arbitrary positive exponent, if

$$G_{ns} = \int_c^d \int_a^b | f(x,y) - T_n(x,y) |^s \, dxdy,$$

and if N is the smallest integer $\geqq 1/s$, the conclusion is that

$$| f(x,y) - T_n(x,y) | \leqq \frac{C_s[(n^2 G_{ns})^{1/s} + \epsilon_n]}{[(b-x)(x-a)(d-y)(y-c)]^{N/2}}$$

at all interior points, with a new C_s which again depends only on s and on the dimensions of the rectangle.

As in the discussion of polynomial approximation just preceding, this result can be interpreted in particular for the middle point of a square, then for the center of a circle, and so ultimately for the interior of an arbitrary closed region.

It is clear that the conclusions that have been formulated above are merely illustrative of a very large number which could be obtained by similar methods.

THE UNIVERSITY OF MINNESOTA,
MINNEAPOLIS, MINN.

---

[18] D. Jackson, *Transactions, loc. cit.*, pp. 246-248, Lemma 6.

ON MULTIPARAMETER EXPANSIONS ASSOCIATED WITH A DIFFERENTIAL SYSTEM AND AUXILIARY CONDITIONS AT SEVERAL POINTS IN EACH VARIABLE.*

By Chester C. Camp.

1. Introduction.

In 1916 C. E. Wilder presented problems in the theory of ordinary differential equations with auxiliary conditions at more than two points.[1] It is the purpose of this paper to extend the theory in two directions.

First by starting with the partial differential equation

$$\sum_{a=1}^{p} L_a(u) + \lambda u = 0 \tag{1}$$

and the auxiliary conditions

$$U_{a\beta}(u) = 0, \tag{2}$$

one is led by Bernoulli's method of solution to what may be called a multiple Wilder system in terms of which one may expand a function $f(x_1, x_2, \cdots, x_p)$ in a multiple Wilder series. Here the coefficient of each new parameter λ_a is unity as in Wilder's differential equation the coefficient of λ is one.

Secondly is considered the system

$$X'_j + [\sum_{i=1}^{p} \lambda_i a_{ji}(x_j)]X_j = 0, \qquad (j = 1, 2, \cdots, p); \tag{3}$$

with the auxiliary conditions

$$X_j(a_j) + c_{j2}X_j(a_{2j}) + \cdots + c_{j,k_j-1}X_j(a_{k_j-1,j}) + c_{j,k_j}X_j(b_j) = 0, \tag{4}$$

where $c_{j,k_j} \neq 0$, $(j = 1, 2, \cdots, p)$.

Here the coefficients of the parameters λ_i will be assumed not to change sign but to maintain their average values in each subinterval. A properly restricted function f will then be expansible in a series of characteristic solutions which will converge almost everywhere to f.

---

* Presented to the American Mathematical Society, in a different form, August 29, 1929. Received by the Editors, August 17, 1937.

[1] *Transactions of the American Mathematical Society*, vol. 18, pp. 415-442 and vol. 19, pp. 157-166.

The system (3) may also be extended to the more general form

$$(3.1) \qquad X'_j + [\sum_{i=1}^{p} \lambda_i a_{ji}(x_j) + g_{ji}(x_j)]X_j = 0, \qquad (j = 1, 2, \cdots, p);$$

which is considerably more general than the system considered in a previous paper by the author.[2]

2. Multiple Wilder system. As in the author's earlier work on Multiple Birkhoff Series[3] one is led from (1), (2) by a simple transformation to a system of p ordinary differential equations of order n_α, $(\alpha = 1, 2, \cdots, p)$ and p boundary systems

$$(5) \qquad L_\alpha(u_\alpha) + \lambda_\alpha u_\alpha = 0, \qquad W_{\alpha\beta}(u_\alpha) = 0, \qquad (\beta = 1, 2, \cdots, n_\alpha).$$

The boundary conditions (2) can be so formulated that, for a particular α, (5) will constitute a Wilder system,[4] where auxiliary conditions are imposed at $k_\alpha > 2$ points. By proceeding as in the article referred to, one may set up multiple series in terms of principal solutions of (5) directly, without reference to (1), (2). In extending the convergence proof from the case of one to p independent variables one may employ the extension of the contour integral method.[5] Although the work is a more arduous piece of routine it can be accomplished in a similar manner[6] without unforeseen difficulties. One may therefore state

THEOREM I. *Given $f(x_1, x_2, \cdots, x_p)$, any real function, which together with its first ν_α*[7] *partial derivatives with respect to x_α (or if $\nu_\alpha < 1$, with its first partial derivative) is continuous in the region (a_α, b_α), $(\alpha = 1, 2, \cdots, p)$; then the multiple Wilder expansion converges to f at every interior point of this region, if the auxiliary conditions are such that for each α certain determinants of the matrix of constants do not vanish and if the $a_{\alpha\beta}$ are such that the*

---

[2] Camp, "An expansion involving p inseparable parameters associated with a partial differential equation," *American Journal of Mathematics*, vol. 50, p. 259, equations (5).

[3] Camp, "Expansions in terms of solutions of partial differential equations," Second Paper, Multiple Birkhoff Series, *Transactions of the American Mathematical Society*, vol. 25, pp. 338-342.

[4] See Wilder's Theorem for restrictions on certain constants in $W_{\alpha\beta}$, Wilder, *loc. cit.*, vol. 18, p. 433.

[5] Camp, Multiple Fourier Series, *Transactions of the American Mathematical Society*, vol. 25, pp. 131, 132.

[6] Cf. the three articles by the author cited above.

[7] See the statement of Wilder's theorem, *loc. cit.*, p. 433.

intervals (a_{2a}, b_a) *and* (a_a, a_{k_a-1}) *are longer than any of the other subintervals of* (a_a, b_a) *included between any two of the points* $a_{a\beta}$, $(\beta = 1, 2, \cdots, k_a)$.

3. **Generalized first order system.** Consider the system (3), (4). If one stipulates that $X_j(x_j)$ shall be continuous, then for each j one may determine a system of conjugate auxiliary conditions by the use of the integral of Lagrange's identity [8] allowing $Y_j(x_j)$, the solution of the system of adjoint differential equations,

$$(6)\qquad -Y'_j + [\sum_{i=1}^{p} \lambda_i a_{ji}(x_j)]Y_j = 0, \qquad (j = 1, 2, \cdots, p);$$

to be discontinuous at each interior auxiliary point a_{ij}, $(i = 2, 3, \cdots, k_j - 1)$. This system as a whole will be unique; it can be written in various forms, one of which is the following:

$$(7)\qquad \begin{cases} c_{j,k_j}Y_j(a_j) + Y_j(b_j) = 0 \\ c_{ji}Y_j(a_j) + Y_j(a_{ij}^-) - Y_j(a_{ij}^+) = 0, \qquad (i = 2, 3, \cdots, k_j - 1). \end{cases}$$

Write a solution of (3) in the form

$$(8)\qquad X_j(x_j) = 1/\exp \sum_{i=1}^{p} \lambda_i A_{ji}(x_j) \quad \text{where} \quad A_{ji}(x_j) = \int_{a_j}^{x_j} a_{ji}(x_j)\,dx_j.$$

A solution of (6) may be written

$$(9)\qquad Y_j(x_j) = K_{ij} \exp \sum_{i=1}^{p} \lambda_i A_{ji}(x_j), \qquad a_{ij} < x_j < a_{i+1,j},$$

$(i = 1, 2, \cdots, k_j - 1)$; where a_{1j} is defined as a_j, and a_{k_j} as b_j. If one takes arbitrarily $K_{1j} = 1$, then the conditions (7) will determine the other K's in succession. The first equation in (7) will determine $K_{k_j-1,j}$ independently, which will be consistent with the other determination provided equation (4) is satisfied. Clearly then the principal parameter values for both systems are the same if they exist. In order to investigate their existence it is expedient to make the following transformations:

$$(10)\quad \nu_j = \sum_{i=1}^{p} \lambda_i A_{ji} \text{ where } A_{ji} = \int_{a_j}^{b_j} a_{ji}(x_j)\,dx_j/(b_j - a_j), \ (j = 1, 2, \cdots, p).$$

The new parameter ν_j is the average value of the coefficient of X_j in (3) over the interval (a_j, b_j). In order to solve backwards we must now assume that the determinant $|A_{ji}|$ does not vanish. If in addition we assume that each $a_{ji}(x_j)$ maintains its average value over each subinterval; i. e.,

---

[8] Wilder, *loc. cit.*, vol. 19, p. 162.

$$(11) \qquad \int_{a_{ij}}^{a_{i+1,j}} a_{jk}(x_j)\,dx_j/(a_{i+1,j}-a_{ij}) = A_{jk},$$
$$(i = 1, 2, 3, \cdots, k_j - 1;\ j = 1, 2, \cdots, p;\ k = 1, 2, \cdots, p);$$

then (4) in view of (8) becomes

$$(12) \qquad W_j(\nu_j) \equiv 1 + \sum_{i=2}^{k_j} c_{ji}\, e^{-(a_{ij}-a_j)\nu_j} = 0, \qquad (j = 1, 2, \cdots, p).$$

If $a_{ji}(x_j)$ is real and integrable then by a theorem given by Langer[9] and due to Wilder and Tamarkin the solutions of (12) lie within a strip $|R(\nu_j)| < K$ and in any rectangular portion of this strip the number of roots $n(R)$ is limited by the relation

$$-n + C(y_2 - y_1) \leqq n(R) \leqq n + C(y_2 - y_1)$$

where y_1, y_2 are two values of the imaginary part of ν_j, $R(\nu_j)$ denotes its real part, and C is the absolute value of the numerically largest coefficient of ν_j in (12).

In case the subintervals are commensurable with the whole interval (a_j, b_j), $W_j(\nu_j)$ reduces to a polynomial of degree p_n in $\exp \alpha\nu_j$ and its zeros are of the form

$$(13) \quad \nu_j = [2m\pi i + \log \xi_k]/\alpha, \quad (k = 1, 2, \cdots, p_n;\ m = 0, \pm 1, \pm 2, \cdots),$$

provided the zeros occur at points for which $\exp \alpha\nu_j = \xi_k$.[10]

In either case when ν_j is uniformly bounded from the roots of (12) then $W_j(\nu_j)$ is uniformly bounded from zero.

The Green's system may be written

$$(14) \quad G \equiv \begin{cases} \prod_{j=1}^{p} [X_j(x_j) Y_j(s_j)/W_j(\nu_j)(b_j - a_j)]/|A_{ji}|, \ s_j < x_j, \\ \qquad\qquad\qquad\qquad\qquad\qquad (j = 1, 2, \cdots, p); \\ \prod_{j=1}^{p} [X_j(x_j) Y_j(s_j)\{1 - W_j(\nu_j)\}/W_j(\nu_j)(b_j - a_j)]/|A_{ji}|, \ s_j > x_j, \\ \qquad\qquad\qquad\qquad\qquad\qquad (j = 1, 2, \cdots, p); \end{cases}$$

where $Y_j(s_j)$ is the discontinuous solution defined by (9), and where the individual factor $1/W_i(\nu_i)$ in the first form of G in (14) is to be replaced by $\{1 - W_i(\nu_i)\}/W_i(\nu_i)$ for every independent variable s_i whenever $s_i > x_i$.

If one wishes to expand a function $f(x_1, x_2, \cdots, x_p)$ as

---

[9] R. E. Langer, "On the zeros of exponential sums and integrals," *Bulletin of the American Mathematical Society*, vol. 37, pp. 213-219. See his references to Wilder and Tamarkin on p. 239.

[10] Cf. Langer, *loc. cit.*, pp. 214-215.

$$(15) \qquad f(x_1, x_2, \cdots, x_p) \sim \sum_{m_j=-\infty}^{\infty} C_{m_1, m_2, \ldots, m_p} \prod X^*_j(x_j) \qquad (j = 1, 2, \cdots, p);$$

where the asterisk denotes a principal solution, then

$$(16) \quad C_{m_1, m_2, \ldots, m_p} = \int_{a_1}^{b_1} \int_{a_2}^{b_2} \cdots \int_{a_p}^{b_p} | a_{ji}(s_j) | \, f(s_1, s_2, \cdots, s_p) \prod_{j=1}^{p} Y^*_j(s_j) \, ds_j / Q$$

where

$$(17) \qquad Q = \int_{a_1}^{b_1} \int_{a_2}^{b_2} \cdots \int_{a_p}^{b_p} | a_{ji}(s_j) | \prod_{j=1}^{n} X^*_j(s_j) Y^*_j(s_j) \, ds_j.$$

If a double star indicates a distinct set of principal parameter values, one has the conjugacy condition

$$(18) \qquad \int_{a_1}^{b_1} \int_{a_2}^{b_2} \cdots \int_{a_p}^{b_p} | a_{ji}(s_j) | \prod_{j=1}^{n} X^*_j(s_j) Y^{**}_j(s_j) \, ds_j = 0.$$

Moreover the residue at a simple set of principal parameter values ν^*_j of the function

$$(19) \qquad \int_{a_1}^{b_1} \int_{a_2}^{b_2} \cdots \int_{a_p}^{b_p} f(s_1, s_2, \cdots, s_p) \ | a_{ji}(s_j) | \ G \prod_{j=1}^{p} ds_j$$

will give the corresponding term of (15) since Q can be shown to take the form

$$(20) \qquad Q = | A_{ji} | \prod_{j=1}^{p} (b_j - a_j) W'_j(\nu^*_j).$$

The convergence proof for the series in (15) is made by the extended contour integral method [11] and the use of the following lemmas.

LEMMA I.

$$\lim_{|z| \to \infty} \frac{1}{2\pi i} \int_{C_z} \frac{1}{W_j(z)} \, \frac{dz}{z} = \tfrac{1}{2}, \qquad c_{j,k_j} \neq 0, \qquad (j = 1, 2, \cdots, p);$$

LEMMA II.

$$\lim_{|z| \to \infty} \frac{1}{2\pi i} \int_{C_z} \frac{| e^{-hz} |}{W_j(z)} \, \frac{dz}{z} = 0, \qquad 0 < h < b_j - a_j;$$

where C_z is a circular contour with center at $z = 0$ uniformly bounded away from zeros of $W_j(z)$. In case $W_j(z)$ contains the factors $1 - e^{\alpha z}$ and $1 - e^{\beta z}$ where α and β are incommensurable it may be necessary to use a sequence of circles C_z of non-uniformly increasing radii as $| z | \to \infty$. The corresponding expansion series will be arranged accordingly in groups of characteristic solutions in order that it may converge properly. With this proviso one may enunciate the

---

[11] Cf. Camp, *American Journal of Mathematics, loc. cit.*, pp. 262 sqq.

THEOREM II. *Let* $f(x_1, x_2, \cdots, x_p)$ *be made up of a finite number of pieces, each real and possessing a continuous partial derivative in each argument in the region* S: $a_j \leqq x_j \leqq b_j$, $(j = 1, 2, \cdots, p)$; *let each* $a_{ji}(x_j)$ *be integrable and either identically zero or of constant sign* $a_j \leqq x_j \leqq b_j$; *let the average value of* $a_{ji}(x_j)$ *over the subinterval* $a_{ij} \leqq x_j \leqq a_{i+1,j}$ *be* A_{ji} *the same for all the subintervals; and let the determinant* $|A_{ji}|$ *be different from zero. Then the expansion* (15) *will converge at any interior point of* S *to the so-called mean value of* f. *If the terms of* (15) *are grouped appropriately the series will converge uniformly to* f *at an interior point of* S *at which* f *is continuous. In the case of a multiple characteristic value it is to be understood that the corresponding term of* (15) *is to be replaced by*

$$\int_{a_1}^{b_1}\int_{a_2}^{b_2}\cdots\int_{a_p}^{b_p} f(s_1, s_2, \cdots, s_p)\,|a_{ji}(s_j)|\,R^*\prod ds_j \tag{21}$$

where R^* *is the residue of the Green's function in* (14).

UNIVERSITY OF NEBRASKA,
LINCOLN, NEBRASKA.

CONCERNING SOME POLYNOMIALS ORTHOGONAL ON A FINITE OR ENUMERABLE SET OF POINTS.*

By MORRIS J. GOTTLIEB.

1. Introduction. Let $x_0 < x_1 < x_2 < \cdots$ be a monotonically increasing sequence of points on the real axis, and $j_0, j_1, j_2, \cdots$ a sequence of positive numbers. Then a system of polynomials $\{p_n(x)\}$ can be defined such that $p_n(x)$ is of the exact degree n, and the orthogonality relations

$$(1.1) \qquad \sum_{\nu=0}^{\infty} j_\nu p_n(x_\nu) p_m(x_\nu) = \begin{cases} 0, n \neq m, \\ 1, n = m, \end{cases} \qquad (n, m = 0, 1, 2, \cdots)$$

hold. These polynomials are uniquely determined except for a factor of ± 1. This follows from the general theory of orthogonal polynomials.

In the "finite" case, in which only a finite number of points $x_0 < x_1 < \cdots < x_N$, and positive values, $j_0, j_1, \cdots, j_N$ are given, a finite system, $p_0(x), p_1(x), \cdots, p_N(x)$, can be defined with the analogous property.

Some special cases occur in the literature. We mention the following:

(a) Tchebychef [10] investigates in detail the "finite" case with $j_0 = j_1 = \cdots = j_N = 1$.

(b) Tchebychef [11] and Ch. Jordan [3] consider the "finite" case of equidistant points with $j_0 = j_1 = \cdots = j_N = 1$. The corresponding polynomials represent a finite analogue of Legendre polynomials. They go over by a proper limiting process into Legendre polynomials.

(c) Krawtchouk [4] considers the finite case, with equidistant points, say $x_\nu = \nu$, and $j_\nu = \binom{N}{\nu} p^\nu q^{N-\nu}$, $(\nu = 0, 1, 2, \cdots, N)$. Here $p > 0$, $q > 0$, $p + q = 1$. He obtains a set of polynomials, which by a suitable limiting process, go over into the Hermite polynomials.

(d) A well known case is that of the Poisson-Charlier polynomials first investigated by Charlier [1]. Here, we have a sequence of equidistant points $x^\nu = \nu$, and $j_\nu = e^{-a} \frac{a^\nu}{\nu!}$, $(\nu = 0, 1, 2, \cdots)$, where $a > 0$.

(e) In this paper, we are concerned with the case of equidistant points, $x_\nu = \nu$ and $j_\nu = e^{-\lambda\nu}$, $(\nu = 0, 1, 2, \cdots; \lambda > 0)$. The relation of the corresponding polynomials to the Laguerre polynomials is similar to that in case (b) to the Legendre polynomials. An indication of this distribution is given by Stieltjes [9], in his classical paper on continued fractions. However, he does not seem to devote any further discussion to these polynomials.

---

* Received October 29, 1937.

The derivation of the usual formal properties of the polynomials (e) presents no difficulties. We mention these formulae in Section 2 without proof and turn our attention in Sections 3, 4 and 5 to the questions of the asymptotic behavior for polynomials of large degree and of the development problem.

In Section 6 some remarks on the cases (c) and (d) are also made.

2. Polynomials of the "Laguerre type." Definition and formal properties. We can show by Abel's transformation (summation by parts) that the polynomials, $l_n(x)$, defined by the formula

$$e^{-\lambda x} l_n(x) = \Delta^n \left\{ e^{-\lambda x} \binom{x}{n} \right\}, \qquad (\lambda > 0), \tag{2.1}$$

satisfy the orthogonality and normalization relations:

$$\sum_{\nu=0}^{\infty} e^{-\lambda\nu} l_n(\nu) l_m(\nu) = \begin{cases} 0, & n \neq m, \\ e^{-n\lambda}(1 - e^{-\lambda})^{-1}, & n = m, \end{cases} \quad (n, m = 0, 1, 2, \cdots). \tag{2.2}$$

This is exactly the case mentioned in § 1, (e).

The highest coefficient in $l_n(x)$ is of the sign $(-1)^n$. We notice that $\sum_{\nu=0}^{\infty} j_\nu = \sum_{\nu=0}^{\infty} e^{-\lambda\nu} = (1 - e^{-\lambda})^{-1}$.

By use of Newton's series, we obtain

$$l_n(x) = e^{-n\lambda} \sum_{\nu=0}^{n} (1 - e^{\lambda})^\nu \binom{n}{\nu} \binom{x}{\nu}. \tag{2.3}$$

The polynomials $l_n(x)$, may be represented by certain Jacobi polynomials. Using the notation of Pólya-Szegö [6, vol. 2, p. 93] we find

$$l_n(x) = P_n^{(0, x-n)}(2e^{-\lambda} - 1). \tag{2.4}$$

If $x = 0, 1, 2, \cdots$, a symmetry property analogous to that of the Poisson-Charlier polynomials[1] is observed from (2.3); namely,

$$e^{n\lambda} l_n(x) = e^{x\lambda} l_x(n), \qquad (x = 0, 1, 2, \cdots). \tag{2.5}$$

In the usual way, the recurrence formula

$$\begin{aligned} (n+1) l_{n+1}(x) - \{(n+1)e^{-\lambda} + n + (e^{-\lambda} - 1)x\} l_n(x) \\ + n e^{-\lambda} l_{n-1}(x) = 0 \end{aligned} \tag{2.6}$$

follows. This is valid for $n = 0, 1, 2, \cdots$, with arbitrary definition of $l_{-1}(x)$.

We notice the following difference equation:

$$\begin{aligned} & e^{-\lambda}(x+2)\Delta^2 l_n(x) \\ & \quad - \{(1 - e^{-\lambda})x + (n-2)e^{-\lambda} - (n-1)\}\Delta l_n(x) \\ & \quad + n(1 - e^{-\lambda}) l_n(x) = 0. \end{aligned} \tag{2.7}$$

For later purposes, the generating function

---

[1] See, for instance, [2].

$$(2.8) \qquad G(x,w) = \sum_{n=0}^{\infty} l_n(x)w^n = (1-w)^x(1-e^{-\lambda}w)^{-x-1}, \quad (|w|<1)$$

is important.

The polynomials, $l_n(x) \equiv l_n(x;\lambda)$ are connected with the Laguerre polynomials by the following limiting relation:

$$(2.9) \qquad \lim_{\lambda\to 0} l_n\left(\frac{x}{\lambda};\lambda\right) = L_n(x),$$

where for the Laguerre polynomials, the notation of Pólya-Szegö [6, vol. 2, p. 94] has been used. This follows immediately from (2.1).

3. Asymptotic formula. By use of the classical method of Darboux we can readily derive an asymptotic formula for $l_n(x)$ which is uniformly valid for large n in a fixed bounded region of the complex x-plane.

The second factor $\phi(w) = (1-e^{-\lambda}w)^{-x-1}$ of the generating function (2.8) is regular in $|w| < e^{\lambda}$; it may be developed in a Taylor's series around $w=1$, so that

$$(3.1) \qquad \begin{cases} G(x,w) = \sum_{\nu=0}^{k} \phi_\nu(1-w)^{x+\nu} + \sum_{\nu=k+1}^{\infty} \phi_\nu(1-w)^{x+\nu} \\ \qquad = S_k(w) + r(w). \end{cases}$$

Here k is a fixed integer, and $\phi_\nu = (-1)^\nu \dfrac{\phi^{(\nu)}(1)}{\nu!}$.

Let now $|x| \leqq \omega$, $(\omega > 0)$; we choose $k > \omega$. Then $r(w)$ is bounded with its first $[k+1-\omega]$ derivatives in $|w| \leqq 1$. Hence, if we write

$$(3.2) \qquad r(w) = \sum_{n=0}^{\infty} c_n w^n,$$

we have

$$(3.3) \qquad c_n = O(n^{-[k+1-\omega]}).$$

If now, k is chosen so that $k > 2\omega + 2$ we find[2]

$$(3.4) \qquad l_n(x) = (-1)^n(1-e^{-\lambda})^{-x-1}\binom{x}{n} + O\{n^{-R(x)-2}\}$$

uniformly in $|x| \leqq \omega$.

The same method leads readily to a complete asymptotic expansion of $l_n(x)$.

4. Location of the zeros. The asymptotic formula (3.4) yields some information concerning the location of the zeros of $l_n(x)$ if n is large. From the general theory of orthogonal polynomials it follows that the zeros of $l_n(x)$ lie on the positive, real axis; more exactly, there are exactly n open intervals of the form, $\nu < x < \nu + 1$, $\nu = 0, 1, 2, \cdots$, each containing one zero in its interior.[3]

---

[2] Here, $R(x)$ denotes the real part of x.

[3] This is generally true in the case of the polynomials defined by (1.1), replacing ν by x_ν. See [7].

From the asymptotic formula (3.4) it is apparent that for any fixed positive integer ν, and an arbitrarily small δ, $0 < \delta < \frac{1}{2}$, an $n(\delta)$ may be found, so that for $n > n(\delta)$, $l_n(x)$ has an odd number of zeros in $\nu - \delta < x < \nu + \delta$. Because of the property mentioned before, there is, then, *exactly one* zero in this interval.

5. Development theorem. E. Schmidt [8] has obtained very complete results on the development of an arbitrary function (defined for non-negative integer values of the argument) in terms of the Poisson-Charlier polynomials. His method can be taken over with slight modifications to discuss the analogous question for our polynomials.

Let $f(x)$ be an arbitrary function defined for $x = 0, 1, 2, \cdots$. We discuss the "Fourier development"

$$f(x) \sim \sum_{n=0}^{\infty} c_n l_n(x) \tag{5.1}$$

where the "Fourier constants," c_n, are defined by

$$c_n = e^{n\lambda}(1 - e^{-\lambda}) \sum_{x=0}^{\infty} e^{-\lambda x} f(x) l_n(x). \quad \text{(See 2.2.)} \tag{5.2}$$

The question is: Under what conditions, regarding $f(x)$, do the expressions c_n have a sense and does the series (51) converge and represent $f(x)$? The first requirement means that the series (5.2) is convergent for each n, $n = 0, 1, 2, \cdots$.

The answer can be stated as follows: A sufficient condition that the development (5.1) exists (i.e., that the constants c_n have a sense) is that the analytic function

$$F(z) = \sum_{\nu=0}^{\infty} f(\nu) z^{\nu}$$

be regular in $|z| < e^{-\lambda}$, and in $\left| z - \frac{1}{1 + e^{\lambda}} \right| < \frac{1}{1 + e^{\lambda}}$, and have bounded derivatives of all orders interior to both circles. If $F(z)$ is regular at $z = e^{-\lambda}$, then this condition is also necessary.

The proof follows closely the discussion of E. Schmidt. The essential idea is the use of the symmetry property (2.5), which is analogous to a similar property of the Poisson-Charlier polynomials.

6. Some remarks concerning Poisson-Charlier polynomials and the polynomials of Krawtchouk.

(a) The Poisson-Charlier polynomials [case (d) of § 1] can be represented in the form

$$(6.1) \qquad p_n(x) = \sum_{\nu=0}^{n} (-1)^{n-\nu} \binom{n}{\nu} \nu!\, a^{-\nu} \binom{x}{\nu}, \qquad (a > 0)$$

cf. Doetsch [2]; they have the following generating function, [2]:

$$(6.2) \qquad H(x, w) = \sum_{n=0}^{\infty} p_n(x) \frac{w^n}{n!} = e^{-w} \left(1 + \frac{w}{a}\right)^x, \qquad (|w| < a).$$

By Darboux's method, the following asymptotic formula for $p_n(x)$, $n \to \infty$, valid uniformly in a fixed region of the complex x-plane is readily obtained from the generating function [4]

$$(6.3) \qquad p_n(x) = e^a a^{-n} n! \binom{x}{n} + a^{-n} n!\, O(n^{-R(x)-2}).$$

As in the case of the polynomials $\{l_n(x)\}$ treated above, it follows that for sufficiently large n, $p_n(x)$ has exactly one zero in each interval $\nu - \delta < x < \nu + \delta$ where ν is fixed, $\nu = 0, 1, 2, \cdots$, and δ is arbitrarily small with $0 < \delta < \frac{1}{2}$.

(b) The Krawtchouk polynomials have the following explicit representation, [4]

$$(6.4) \qquad k_n(x) = \sum_{\nu=0}^{n} (-1)^{n-\nu} \binom{N-x}{n-\nu} \binom{x}{\nu} p^{n-\nu} q^{\nu}.$$

Here, $p > 0$, $q > 0$, and $p + q = 1$. The polynomials defined by this formula have a sense also for $n > N$, although they then cease to satisfy the normalization conditions, since they vanish for $x = 0, 1, 2, \cdots, N$.

For these polynomials also, an asymptotic formula can be obtained by Darboux's method, valid uniformly for large n, in a fixed region of the complex x-plane. We start out from the generating function, [4],

$$(6.5) \qquad K(x, w) = \sum_{n=0}^{\infty} k_n(x) w^n = (1 + qw)^x (1 - pw)^{N-x}, \quad \left[|w| < \min\left(\frac{1}{p}, \frac{1}{q}\right) \right].$$

It is necessary to distinguish three different cases according to different values of the constants p and q. In case $p < q$, the asymptotic formula is

$$(6.6) \qquad k_n(x) = q^n \left\{ q^{-(N-x)} \binom{x}{n} + O(n^{-R(x)-2}) \right\}.$$

By interchanging p and q, x and $N - x$, w and $-w$, the asymptotic formula for $p > q$ is obtained from the previous case:

---

[4] Doetsch [2] mentions the possibility of obtaining the asymptotic formula in this way, without discussing it further. Cf. also the preliminary communication of Obrechkoff [5].

$$k_n(x) = p^n \left\{ (-1)^n p^{-x} \binom{N-x}{n} + O(n^{-[N-R(x)]-2}) \right\}. \tag{6.7}$$

In the case $p = q = \frac{1}{2}$, the asymptotic formula obtained by Darboux's method is

$$k_n(x) = 2^{-n} \left\{ 2^{N-x} \binom{x}{n} + (-1)^n 2^x \binom{N-x}{n} + O(n^{-\xi-2}) \right\}, \tag{6.8}$$

where $\xi = \min \{R(x), N - R(x)\}$.

All three formulae are valid uniformly in a fixed region of the complex x-plane. The same method leads readily to a complete asymptotic expansion of $k_n(x)$.

WASHINGTON UNIVERSITY,
ST. LOUIS, MO.

BIBLIOGRAPHY.

1. C. Charlier, "Über das Fehlergesetz.—Über die Darstellung willkürlicher Funktionen," *Arkiv för Matematik, Astronomy och Fysik*, vol. 2 (1905-1906), no. 8 and 20.

2. G. Doetsch, "Die in der Statistik seltener Ereignisse auftretenden Charlierschen Polynome und eine damit zusammenhängende Differentialdifferenzengleichung," *Mathematische Annalen*, vol. 109 (1933), pp. 257-266.

3. Ch. Jordan, "Sur une série de polynomes dont chaque somme partielle représente le meilleure approximation d'un degré donné suivant la méthode des moindre carrés," *Proceedings of the London Mathematical Society*, ser. 2, vol. 20 (1921), pp. 297-325.

4. M. Krawtchouk, "Sur une généralization des polynomes d'Hermite," *Comptes Rendus de l'Académie des Sciences*, Paris, vol. 189 (1929), pp. 620-621.

5. N. Obrechkoff, "Sur une classe de polynomes," *Comptes Rendus du Congrès International des Mathématiciens*, Oslo, 1936, vol. 2 (1937), pp. 115-116.

6. G. Pólya and G. Szegö, *Aufgaben und Lehrsätze aus der Analysis*, I, II, Berlin, 1925.

7. T. Popoviciu, "Sur la distribution des zéros de certains polynomes minimisants," *Bulletin Sect. Sci. Acad. Roum.*, vol. 16 (1934), pp. 214-217.

8. E. Schmidt, "Über die Charlier-Jordansche Entwicklung einer willkürlichen Funktion nach der Poissonschen Funktion und ihren Ableitungen," *Zeitschrift für angewandte Mathematik und Mechanik*, vol. 13 (1933), pp. 139-142.

9. T. J. Stieltjes, "Recherches sur les fractions continues," *Ann. Fac. Sci. Toulouse*, vol. 8 (1894), pp. 1-122; vol. 9 (1895), pp. 1-47. Cf. *Oeuvres complètes*, vol. 2, pp. 546-547.

10. P. L. Tchebychef, "Sur une formule d'analyse," *Bull. Phys. Math. de l'Académie Impériale des Sciences de St. Petersbourg*, vol. 13 (1854), pp. 210-211; "Sur les fractions continues," *Journal de Mathématiques*, ser. 2, vol. 3 (1855), pp. 289-323.

11. P. L. Tchebychef, "Sur l'interpolation par la méthode des moindres carrés," *Mémoirs de l'Acad. Imp. des Sciences de St. Pétersbourg*, ser. 7, vol. 1 (1859), pp. 1-24.

ON THE FOURIER-STIELTJES TRANSFORM OF A SINGULAR FUNCTION.*

By Philip Hartman and Richard Kershner.

It is a well known consequence of the Riemann-Lebesgue lemma that the Fourier-Stieltjes transform,

$$(1)\qquad L(t;\sigma)=\int_{-\infty}^{\infty}\exp(itx)\,d\sigma(x),$$

of an absolutely continuous distribution function $\sigma(x)$ satisfies

$$(2)\qquad L(t;\sigma)=o(1),\qquad t\rightarrow\pm\infty.$$

On the other hand, if $\sigma(x)$ is any distribution function for which

$$L(t;\sigma)=O(|\,t\,|^{-1-\epsilon}),\qquad t\rightarrow\pm\infty,$$

for some $\epsilon>0$, then $\sigma(x)$ is necessarily absolutely continuous.

If $\sigma(x)$ is a purely singular distribution function, then (2) need not hold.[1] The first example to the effect that (2) may hold when t runs through integral multiples of 2π, even when $\sigma(x)$ is purely singular, was given by Menchoff[2] in the analogous case of Fourier-Stieltjes coefficients of a periodic function. A somewhat simpler example of a purely singular distribution function $\sigma(x)$ for which not only (2) held but

$$(3)\qquad L(t;\sigma)=O(\log^{-\gamma}|\,t\,|),\qquad t\rightarrow\pm\infty$$

for a certain positive γ, was given by one of the authors.[3] More recently, Littlewood[4] has given a very complicated example of a purely singular distribution function $\sigma(x)$ for which

---

* Received August 17, 1937.

[1] An example to this effect is the Cantor function.

[2] D. Menchoff, "Sur l'unicité du développement trigonométrique," *Comptes Rendus*, vol. 163 (1916), pp. 433-436.

[3] R. Kershner, "On singular Fourier-Stieltjes transforms," *American Journal of Mathematics*, vol. 58 (1936), pp. 450-452.

[4] J. E. Littlewood, "On the Fourier coefficients of functions of bounded variation," *Quarterly Journal of Mathematics, Oxford Series*, vol. 7 (1936), pp. 219-226. Actually, Littlewood constructed a function $f(x)$ of period 1, which is continuous for all x and of bounded variation in $0\leqq x\leqq 1$, such that if a_n, $n=0,\pm 1,\cdots$ are the Fourier coefficients of $f(x)$, then $a_n=O(|\,n\,|^{-1-\epsilon})$. This function was of the type $f_k(x)$

$$(4)\qquad L(2\pi n;\sigma)=O(|\,n\,|^{-c}),\qquad n\to\pm\infty\qquad(n=0,\pm1,\cdots)$$

for some positive c. In all the above cases, the distribution functions treated were not only purely singular but, in fact, almost everywhere constant, i. e., in the language of distribution functions, their spectra were zero sets.

The object of the present note is to show that (2) and, in fact, (3) may hold in the case of a purely singular distribution function the spectrum of which is an interval. It may be mentioned that, for the example to be given, it is certain that (4) does not hold.

Let $\sigma_n=\sigma_n(x)$ denote the distribution function

$$\begin{array}{lll}\sigma_n(x)=0; & -\infty<x\leqq 0; & \\ \sigma_n(x)=\frac{1}{2}(1+n^{-\frac{1}{2}}); & 0<x\leqq(\frac{1}{2})^n; & (n=1,2,\cdots); \\ \sigma_n(x)=1; & (\frac{1}{2})^n<x<+\infty. & \end{array}$$

Let $\tau(x)$ denote the infinite convolution

$$\tau(x)=\sigma_1*\sigma_2*\cdots.$$

This Poisson convolution is absolutely convergent[5] and represents a continuous[6] purely singular[6] distribution function. Furthermore, the spectrum of $\tau(x)$ is the interval $0\leqq x\leqq 1$, so that in this interval $\tau(x)$ is strictly increasing.

With the above notations it will be shown that the *purely singular distribution function* $\tau(x)$, *whose spectrum is an interval, has a Fourier transform* $L(t;\tau)$ *which satisfies*

$$(7)\qquad L(t;\tau)=O(\log^{-\frac{1}{2}}|\,t\,|),\qquad t\to\pm\infty$$

and also

$$(8)\qquad L(t;\tau)=\Omega(\log^{-\frac{1}{2}}|\,t\,|),\qquad t\to\pm\infty.$$

Obviously,

$$(9)\qquad L(t;\sigma_n)=\tfrac{1}{2}(1+n^{-\frac{1}{2}})+\tfrac{1}{2}(1-n^{-\frac{1}{2}})\exp(i2^{-n}t).$$

Hence, by the multiplication rule of Fourier transforms,

$$(10)\qquad L(t;\tau)=\prod_{n=1}^{\infty}\left[\tfrac{1}{2}(1+n^{-\frac{1}{2}})+\tfrac{1}{2}(1-n^{-\frac{1}{2}})\exp(i2^{-n}t)\right],$$

$=\sigma(x)-x$, $0\leqq x\leqq 1$, where $\sigma(x)$ is a purely singular monotone function such that $\sigma(0)=0$, $\sigma(1)=1$. Now, $\sigma(x)$ may be considered to be a distribution function if one places $\sigma(x)=0$ for $-\infty<x\leqq 0$ and $\sigma(x)=1$ for $1\leqq x<+\infty$. It is clear that $a_n=(2\pi in)^{-1}L(-2\pi n;\sigma)$.

[5] B. Jessen and A. Wintner, "Distribution functions and the Riemann zeta function," *Transactions of the American Mathematical Society*, vol. 38 (1935), pp. 48-88.

[6] P. Hartman and R. Kershner, "On the structure of monotone functions," *American Journal of Mathematics*, vol. 59 (1937), pp. 809-822.

so that

$$|L(t;\tau)|^2 = \prod_{n=1}^{\infty} [\tfrac{1}{2}(1+n^{-1}) + \tfrac{1}{2}(1-n^{-1}) \cos 2^{-n}t]$$

or

$$|L(t;\tau)|^2 = \prod_{n=1}^{\infty} [n^{-1} \sin^2 2^{-n-1}t + \cos^2 2^{-n-1}t]. \tag{11}$$

Now let λ denote the set of all points $t > 0$ which are within a distance $\pi/6$ of an integral multiple of π, so that there exists a positive $\delta < 1$ such that

$$|n^{-1} \sin^2 t + \cos^2 t| < \delta < 1, \ n > 1, \text{ if } t \not\subset \lambda. \tag{12}$$

Also let μ_n denote the set of all points $t > 0$ which are within a distance $2^{-n}\pi/6$ of an odd multiple of $\pi/2$, so that

$$|\cos t| \leqq 2^{-n-1}, \text{ if } t \subset \mu_n. \tag{13}$$

Now let $t \geqq 2\pi$ be fixed and let m $(\geqq 1)$ be the unique integer such that

$$2^m\pi \leqq t < 2^{m+1}\pi. \tag{14}$$

Suppose that k $(0 \leqq k \leqq m)$ of the m values $1, 2, \cdots, m$ of n are such that $2^{-n-1}t$ is in λ. Suppose that these k values of n consist of j $(1 \leqq j \leqq k)$ sets g_i, $(i = 1, 2, \cdots, j)$, each composed of l_i $(k \geqq l_i \geqq 1)$ successive integers, so that

$$\sum_{i=1}^{j} l_i = k. \tag{15}$$

Let the groups g_i be ordered in such a way that the integers contained in g_i are less than those contained in g_{i+1} $(i = 1, 2, \cdots, j-1)$. Thus the integers n_i immediately following the group of successive integers g_i satisfies

$$n_i \geqq \sum_{r=1}^{i} l_r + 1 \tag{16}$$

and

$$2^{-n_i-1}t \subset \mu_{l_i}. \tag{17}$$

Thus each group g_i of integers consisting of indices of large factors of the product (11) proves the existence of an integer n_i such that, by (13) and (17),

$$n_i^{-1} \sin^2 2^{-n_i-1}t + \cos^2 2^{-n_i-1}t < n_i^{-1} + 2^{-2(l_i+1)}. \tag{18}$$

This inequality (18) is used to majorize j of the factors of the product (11) with indices among the $m-k$ values of n $(n = 1, 2, \cdots, m)$ for which $2^{-n-1}t$ is not in λ. The remaining $m-k-j$ $(\geqq 0)$ factors of this set are majorized by (12). Finally, all other factors are replaced by 1. Thus

$$|L(t;\tau)|^2 \leqq \delta^{m-k-j} \prod_{i=1}^{j} (n_i^{-1} + 2^{-2(l_i+1)})$$

or, using (16),

$$(19) \qquad |L(t;\tau)|^2 \leqq \delta^{m-k-j} \prod_{i=1}^{j} [(1 + \sum_{r=1}^{i} l_r)^{-1} + 2^{-2(1+l_i)}].$$

Now

$$(20) \quad [(1+l_1)^{-1} + 2^{-2(1+l_1)}] [(1+l_1+l_2)^{-1} + 2^{-2(1+l_2)}] < \tfrac{9}{16}(1+l_1+l_2)^{-1} + 2^{-2(1+l_2)}(1+l_1)^{-1} + 2^{-2(1+l_1+l_2)}.$$

Also, clearly,

$$(21) \qquad 2^{-2(1+l_2)}(1+l_1)^{-1} < \tfrac{1}{4}(1+l_1+l_2)^{-1} < \tfrac{7}{16}(1+l_1+l_2)^{-1},$$

so that, by (20), (21),

$$(22) \qquad [(1+l_1)^{-1} + 2^{-2(1+l_1)}] [(1+l_1+l_2)^{-1} + 2^{-2(1+l_2)}] < (1+l_1+l_2)^{-1} + 2^{-2(1+l_1+l_2)}.$$

By repeated use of the appraisal (22), formula (19) becomes

$$|L(t;\tau)|^2 < \delta^{m-k-j}[(1 + \sum_{i=1}^{j} l_i)^{-1} + 2^{-2(1+\sum_{i=1}^{j} l_i)}],$$

or

$$(23) \qquad |L(t;\tau)|^2 < \delta^{m-k-j} 2/k, \qquad m-k-j \geqq 0,\ j \leqq k.$$

If $k < m/2$, then, by (23),

$$|L(t;\tau)|^2 < [4(m\delta^m)/(2k\delta^{2k})](1/m) = O(1/m).$$

On the other hand, if $k \geqq m/2$,

$$|L(t;\tau)|^2 < \delta^0 \cdot 2 \cdot 2/m = O(1/m).$$

Thus, in either case, (7) follows from the definition (14) of the integer m.

To prove (8) let $t = 2^m\pi$. Then, by (11),

$$|L(2^m\pi;\tau)|^2 = m^{-1} \prod_{n=m+1}^{\infty} [n^{-1} \sin^2 2^{-n+m-1}\pi + \cos^2 2^{-n+m-1}\pi],$$

so that, neglecting the first term in each factor

$$|L(2^m\pi;\tau)|^2 > Cm^{-1} \qquad [C = \prod_{n=1}^{\infty} \cos^2(2^{-n}\pi/2) > 0].$$

This completes the proof of the italicized statement above.

THE JOHNS HOPKINS UNIVERSITY,
THE UNIVERSITY OF WISCONSIN.

LIOUVILLE SYSTEMS AND ALMOST PERIODIC FUNCTIONS.*

By Aurel Wintner.

Since the Staude-Stäckel theory of conditionally periodic systems has been made standard through Charlier's textbook on celestial mechanics (1902; also 1927), and since Charlier's presentation follows closely that of Stäckel's paper of 1891, it is usually overlooked that, as pointed out by Stäckel himself (1905) and by Hadamard (1911), this simple and general theory leads to several difficulties. These difficulties of the standard theory are quite serious, since the objections of Stäckel and Hadamard, until they are met, prohibit any application of the Staude-Stäckel theory. In fact, the difficulty observed by Stäckel [7] is this: while the standard theory assumes that a certain Jacobian is distinct from zero in the whole domain under consideration, it turns out that this condition is necessarily violated even in the simplest possible cases (e. g., in the case of geodesics on a surface of Liouville type). The difficulty pointed out by Hadamard [5], being not of a local nature, is still more fundamental and concerns the usual introduction of the uniformizing variables into the (real) inversion problem of the Abel-Jacobi type; the possibility of defining these uniformizing variables, i. e., the problem of the monodromy group, being considered as settled by the local non-vanishing of the Jacobian involved.

In order to legitimize the procedure of the standard theory, Hadamard [5] has indicated direct topological considerations which, by proceeding from case to case, enable one to verify that the difficulties mentioned above can be disposed of. Furthermore, Stäckel [7] has shown by a detailed analysis of the uniformization belonging to bounded geodesics on surfaces of the Liouville type that in this particular case the local vanishing of the Jacobian, which is a necessity, cannot influence the correctness of the final result. Needless to say, such direct discussions can become quite involved, if one wants to take care of all the possible cases.

It will be shown in what follows that a straightforward and rigorous treatment of an extended class of separable systems can be based on Bohr's theory of almost periodic functions in such a way that no explicit discussions are needed when the theory is applied to concrete cases. While the class of systems to be considered does not include the most general separable system, it contains practically all of the classical integrable problems in the dynamics

---

* Received December 23, 1937.

of particles, since it consists of the systems usually associated with the name Liouville. (Actually, there is no restriction at all in the fundamental case of $n = 2$ degrees of freedom, since it is known that in this case the most general separable system is a Liouville system, if one allows a coördinate transformation of the trivial type

$$\bar{x} = F_1(x) + F_2(y), \qquad \bar{y} = G_1(x) + G_2(y)$$

and assumes, as usual, that the system is reversible; cf. Stäckel [7], where further references are given.) That the problem of uniformization can be treated directly by means of the theory of almost periodic functions, is due to the fact that, in the case of Liouville systems, the inversion problem of the Abel-Jacobi type reduces to that of a synchronization of n given uniformizing parameters along n one-dimensional closed manifolds whose product space is the n-dimensional torus of the separable system.

Denoting by dots differentiations with respect to t, the general conservative Lagrangian function of the non-relativistic dynamical type and of n degrees of freedom is

$$(1) \qquad L \equiv L(x, \dot{x}) = \tfrac{1}{2} \Sigma \Sigma\, g_{ik} \dot{x}_i \dot{x}_k + \Sigma f_i \dot{x}_i + e, \qquad (\Sigma = \sum_{1}^{n}),$$

where $g_{ik} = g_{ki}$, f_i, e are functions of $x = (x_1, \cdots, x_n)$ which have continuous partial derivatives of the second order in the x-domain under consideration, while the matrix $\| g_{ik} \|$ is positive definite in this domain. The reversible case is characterized by the identical vanishing of the n functions $f_i(x)$. Whether this condition is or is not satisfied, the Lagrangian equations have the energy integral

$$(2) \qquad \tfrac{1}{2} \Sigma \Sigma\, g_{ik}(x) \dot{x}_i \dot{x}_k - e(x) = h \qquad (h = \text{const.}).$$

In order to extend Liouville's type to the irreversible case, suppose that the $\frac{1}{2}n(n+1) + n + 1$ functions $g_{ik} = g_{ki}$, f_i, e of the n variables $x_1, \cdots, x_n$ can be expressed in terms of n sets of 4 functions

$$(3) \qquad g_i = g_i(x_i), \quad f_i = f_i(x_i), \quad e_i = e_i(x_i), \quad d_i = d_i(x_i)$$

of the single variable x_i, where $i = 1, \cdots, n$, in such a way that the Lagrangian function (1) becomes †

---

† The content of the assumption (4) becomes clearer if one replaces the velocities $\dot{x}_i$ by the momenta $y_i = \partial L / \partial \dot{x}_i$ and L by the Hamiltonian function H. In fact, it is easily verified that (1) has the particular structure (4), (5) if and only if H is representable in terms of n pairs H_i^1, H_i^2 of functions of the two variables x_i, y_i in the symmetrical form

$$H(x_1, \cdots, x_n, y_1, \cdots, y_n) = \Sigma H_i^1(x_i, y_i) : \Sigma H_i^2(x_i, y_i).$$

$$(4)\qquad L \equiv L(x, \dot{x}) = \tfrac{1}{2} r(x_1, \cdots, x_n) \Sigma g_i(x_i) \dot{x}_i^2 + \Sigma f_i(x_i) \dot{x}_i + \Sigma e_i(x_i)/r(x_1, \cdots, x_n),$$

where

$$(5)\qquad r = r(x_1, \cdots, x_n) = \Sigma d_i(x_i).$$

Then the positive definite matrix $\| g_{ik} \|$ is the diagonal matrix formed by the products $r(x_1, \cdots, x_n) g_i(x_i)$, and so

$$(6_1)\quad r(x_1, \cdots, x_n) > 0; \qquad (6_2)\quad g_i(x_i) > 0, \qquad (i = 1, \cdots, n),$$

if the notations are chosen in a suitable way.

According to (6_1), one can introduce along any given solution path $x = x(t)$ a new time variable, t^*, by placing

$$(7)\quad t^* \equiv t^*(t) = \int^{t} d\bar{t}/r\big(x_1(\bar{t}), \cdots, x_n(\bar{t})\big), \text{ i. e., } \dot{t}^* = 1/r,\ t' = r,$$

where the prime denotes differentiation with respect to t^*. If the energy constant of $x = x(t)$ is h, put

$$(8)\qquad L^* = L^*(x, x'; h) = \tfrac{1}{2} \Sigma g_i x'^2_i + \Sigma f_i x'_i + \Sigma e_i + h \Sigma d_i.$$

The energy integral of the Lagrangian equations belonging to (8) is

$$(9)\qquad \tfrac{1}{2} \Sigma g_i(x_i) x'^2_i - \{\Sigma e_i(x_i) + h \Sigma d_i(x_i)\} = h^* \qquad (h^* = \text{const.}).$$

Now, those solutions $x = x(t)$ of the Lagrangian equations belonging to (4) which have the energy h are, in virtue of the time transformation (7) or its inverse

$$(10)\qquad t \equiv t(t^*) = \int^{t^*} r\big(x_1(\bar{t}^*), \cdots, x_n(\bar{t}^*)\big) d\bar{t}^*,$$

identical with those solutions $x = x(t^*)$ of the Lagrangian equations belonging to (8) which have the energy

$$(11)\qquad h^* = 0.$$

In order to see this, it is sufficient to observe that the Lagrangian function of the Maupertuis principle belonging to (1) is

$$\Lambda(x, \dot{x}; h) = [(e + h)(\Sigma\Sigma g_{ik} \dot{x}_i \dot{x}_k)]^{\frac{1}{2}} + \Sigma f_i x_i,$$

where, in view of (4), (5) and (6_1), (6_2),

$$[\quad]^{\frac{1}{2}} = [(r^{-1} \Sigma e_i + h)(r \Sigma g_i \dot{x}_i^2)]^{\frac{1}{2}} = [(\Sigma e_i + h \Sigma d_i)(\Sigma g_i \dot{x}_i^2)]^{\frac{1}{2}};$$

while $\Lambda(x, q\dot{x}; h) = q\Lambda(x, \dot{x}; h)$ for every $q > 0$, and so, in particular, for $q = r$ or $q = r^{-1}$.

The Lagrangian function (4) seems to be more general than a usual Lagrangian function of the Liouville type, since in (4) the terms linear in the velocities are not missing. Actually, one can omit the term $\Sigma f_i(x_i)\dot{x}_i$ without changing the Lagrangian equations, since $\Sigma f_i(x_i)dx_i$ is a complete differential. Thus, in contrast with the integrable irreversible dynamical systems investigated for $n=2$ by Birkhoff ([2], pp. 206-210), one can put $f_i(x_i)\equiv 0$ without loss of generality. Then the function (8) can be written as ΣL_i, where

$$(12_1)\quad L_i=\tfrac{1}{2}g_i(x_i)x'^2_i+U_i(x_i;h);\qquad (12_2)\quad U_i=e_i(x_i)+hd_i(x_i).$$

Hence the system of n Lagrangian equations $[L^*]_{x_i}=0$ splits into the n Lagrangian equations $[L_i]_{x_i}=0$ each of which has, in view of (12_1), a single degree of freedom, and, correspondingly, an energy integral

$$(13)\qquad \tfrac{1}{2}g_i(x_i)x'^2_i-U_i(x_i;h)=h_i\qquad (h_i=\text{const.}).$$

However, the choice of the n integration constants h_i is restricted by the condition

$$(14)\qquad \Sigma h_i=0,$$

this condition being identical with (9), since $\Sigma h_i=h^*$ in view of (13), (12_2) and (9).

According to (6_2), one can write (13) in the form

$$(15_1)\quad x'^2_i=F_i(x_i;h_i,h);\qquad (15_2)\quad F_i=\frac{U_i(x_i;h)+h_i}{\tfrac{1}{2}g_i(x_i)}.$$

And $x_i=x_i(t^*)$ follows from (15_1), where $x'_i=dx_i/dt^*$, by the inversion of a quadrature. Since $x_i(t^*-t_i)$ is, for every solution $x_i(t^*)$ of $[L_i]_{x_i}=0$ and for every constant t_i, a solution of $[L_i]_{x_i}=0$, the integration constant introduced by the quadrature can be omitted, so that the motion is, in the main, determined for every i by the energy constant h_i alone. In this sense, it is permissible to denote the general solution by

$$(16)\qquad x_i=x_i(t^*;h_i,h),$$

where one can add to t^* different arbitrary constants for different values of i.

Suppose, for simplicity, that the x_i-space is the whole space $-\infty<x_i<+\infty$, where $i=1,\cdots,n$, and assume that if $F_i(x_i;h_i,h)>0$ for large $x_i>0$ or large $-x_i>0$, then

$$\int^{\pm\infty}F_i(\bar{x}_i;h_i,h)^{-\frac{1}{2}}\,d\bar{x}_i=\pm\infty\quad(\text{e.g., } |F_i|<\text{const. } x_i^2).$$

Since, from (15_1),

$$(17)\qquad t^* = \int^{x_i} \{F_i(\bar{x}_i; h_i, h)\}^{-\frac{1}{2}}\, d\bar{x}_i,$$

the greatest lower bound, α_i, and the least upper bound, β_i, of (16) for $-\infty < t^* < +\infty$ are two *subsequent* roots

$$(18)\qquad x_i = \alpha_i \equiv \alpha_i(h_i, h), \qquad x_i = \beta_i \equiv \beta_i(h_i, h)$$

of the equation

$$(19)\qquad F_i(x_i; h_i, h) = 0$$

of Hill's manifold of zero velocity [cf. (15_1)], where it is understood that $\alpha_i = -\infty$ and/or $\beta_i = +\infty$ in case such roots do not exist, and that $\alpha_i = \beta_i$ if and only if (16) is independent of t^*. If (16) is for some t^* between two subsequent distinct multiple roots (18) of the equation (19), then the integrand of (17) becomes infinite in a non-integrable order ($\geqq 1$) when $x_i \rightarrow \alpha_i + 0$ (or $x_i \rightarrow \beta_i - 0$), and so (16) tends to α_i (or β_i) either when $t^* \rightarrow -\infty$ or when $t^* \rightarrow +\infty$ or when $t^* \rightarrow \pm\infty$. These asymptotic cases will be excluded in what follows. Then there exist two subsequent simple roots (18) of (19) such that α_i is the minimum and β_i the maximum of (16) for $-\infty < t^* < +\infty$, the integrand of (17) becoming infinite at $x_i = \alpha_i$ and $x_i = \beta_i$ in the integrable order $\frac{1}{2}$. Thus, from (18) and (19),

$$(20)\qquad F_i(x_i; h_i, h) = (\beta_i - x_i)(x_i - \alpha_i) G_i, \qquad (\alpha_i < \beta_i),$$

where, by (15_1),

$$(21)\qquad G_i(x_i; h_i, h) > 0 \quad \text{for} \quad \alpha_i \leqq x_i \leqq \beta_i,$$

and G_i remains continuous at $x_i = \alpha_i, \beta_i$. Let h and h_i be fixed for a fixed i, so that (16), (20), (21) reduce to $x_i(t^*)$, $F_i(x_i)$, $G_i(x_i)$ respectively.

It is easily inferred from (17), (20), (21) that (16) is periodic with the primitive period

$$(22)\qquad \tau_i = 2\int_{\alpha_i}^{\beta_i} \{F_i(\bar{x}_i)\}^{-\frac{1}{2}}\, d\bar{x}_i.$$

One can uniformize the relation (17) between t^* and x_i in the usual manner (Abel, Hill, Weierstrass) in terms of a uniformizing time variable u_i which varies with t^* from $-\infty$ to $+\infty$ and reduces to the eccentric anomaly in case of Kepler's motion (where $n = 1$, $x_i = x_1 =$ radius vector, $t^* =$ time and the degree of freedom is reduced to $n = 1$ by considering the true anomaly as a coördinate which is ignorable in the sense of Routh). This uniformization of (17) is

$$(23_1)\quad x_i = \tfrac{1}{2}(\beta_i + \alpha_i) - \tfrac{1}{2}(\beta_i - \alpha_i)\cos u_i; \qquad (23_2)\quad t^* = \tfrac{1}{2\pi}\tau_i u_i + P_i(u_i),$$

where τ_i is the fixed period (22) and the function $P_i(u)$ is such that

$$(24_1) \quad dt^*/du_i > 0; \qquad (24_2) \quad P_i(u_i + 2\pi) = P_i(u);$$

(it is essential that in (24_1) the sign of equality, which would be compatible with a strictly increasing function (23_2) of u_i, is excluded for every u_i). For let the derivative (24_1), considered as a function of u_i, be defined as the positive square root of the function which one obtains by substituting the function (23_1) of u_i into the positive continuous function (21) of x_i. Then the function (24_1) of u_i has the period 2π and over this period a positive mean value, say γ_i. Hence, $t^* = \gamma_i u_i + P_i(u_i)$ holds for a function $P_i(u_i)$ which satisfies (24_2). In order to see that this implies the uniformization of the relation (17) in the form (23_1), (23_2), where $-\infty < u_i < +\infty$, it is sufficient to compare (20) with the representation (22) of the period of (16), and to observe that (23_1) is equivalent to $(dx_i/du_i)^2 = (\beta_i - x_i)(x_i - \alpha_i)$, while $\min x_i(t^*) = \alpha_i$, $\max x_i(t^*) = \beta_i$.

The n periods τ_i with respect to t^* are, in view of (18), (20) and (22), continuous functions of the integration constants h_i, h and are, therefore, incommensurable in general. Hence the uniformization of the solutions $x_i = x_i(t)$ of the unseparated system of the original Lagrangian equations $[L]_{x_i} = 0$ is not an elementary task. For in order to obtain the n functions $x_i = x_i(t)$, where $-\infty < t < +\infty$, one has to eliminate the $n+1$ intermediary time variables u_i, t^* between the $2n+1$ equations (23_1), (23_2), (10). This non-local elimination will be carried out by using a theorem on almost periodic functions, almost periodicity being meant in the original sense of Bohr.

If $\phi = \phi(t^*)$, $-\infty < t^* < +\infty$, is a real almost periodic function which has a derivative $\phi'(t^*) > -1$ for every t^*, then, $\phi(t^*)$ being bounded, the function $t = t^* + \phi(t^*)$ of t^* varies with t^* from $-\infty$ to $+\infty$ in a strictly increasing manner. This does not imply that the almost periodic function $t - t^*$ of t^* is an almost periodic function of t. If, however, the assumption $-1 < \phi'(t^*)$ is replaced by the sharper assumption that there exists a ϑ for which

$$(25) \quad -1 < -\vartheta \leqq \phi'(t^*), \text{ where } \vartheta = \text{const. for } -\infty < t^* < +\infty,$$

then, on placing

$$(26) \qquad t^* = t + \psi(t), \text{ where } t = t^* + \phi(t^*),$$

the almost periodicity of $\phi(t^*)$ implies the almost periodicity of $\psi(t)$; furthermore, the moduls determined by the frequencies of $\phi(t^*)$ and $\psi(t)$ are identical (cf. Bohr [3], where it is assumed that $|\phi'(t^*)| \leqq \vartheta < 1$, while actually (25) is sufficient; cf. Bohr and Jessen [4]).

Returning to the uniformization (23_1), (23_2), where τ_i is a positive constant and the correspondence between $-\infty < u_i < +\infty$ and $-\infty < t^* < +\infty$ is topological, it is clear that the inverse of the function (23_2) can be written in the form

(27_1) $u_i = \sigma_i t^* + Q_i(t^*)$; (27_2) $\sigma_i = 2\pi/\tau_i$,

where, in view of (24_1) and (24_2),

(28_1) $u'_i > 0$, i. e., $Q'_i(t^*) > -\sigma_i$; (28_2) $Q_i(t^* + \tau_i) = Q_i(t^*)$.

Since $x_i(t^* + \tau_i) = x_i(t^*)$, and since the functions (3) depend only on x_i, one has

(29) $d_i(t^* + \tau_i) = d_i(t^*)$, where $d_i(t^*) = d_i(x_i(t^*))$.

Hence (5) is a sum of n functions of t^* which have the periods $\tau_1, \cdots, \tau_n$ respectively, so that

(30) $$r(t^*) = \Sigma d_i(t^*)$$

is almost periodic. From (10) and (30) one has

(31_1) $t \equiv t(t^*) = \int^{t^*} r(\bar{t}^*) d\bar{t}^* = \Sigma s_i(t^*)$; (31_2) $s_i(t^*) = \int^{t^*} d_i(\bar{t}^*) d\bar{t}^*$.

Let, as usual, $M\{f\}$ denote the mean value of an almost periodic (e. g., continuous periodic) function f, i. e., the constant term in the Fourier expansion of f. It is clear from (31_2) and (29) that the difference of $s_i(t^*)$ and $t^* M\{d_i\}$ has the period τ_i. It follows, therefore, from (31_1), (30) that, if $\mu = \Sigma M\{d_i\}$, one has

(32_1) $t = \mu t^* + v(t^*)$; (32_2) $v(t^*) = \Sigma p_i(t^*)$,

where

(33_1) $p_i(t^* + \tau_i) = p_i(t^*) \equiv s_i(t^*) - t^* M\{s_i\}$; (33_2) $\mu = M\{r\}$.

Since (5) is, by (6_1), everywhere positive, and since (5) is a continuous function of the position in the space $x = (x_1, \cdots, x_n)$, it has on every bounded subset of this space a non-vanishing greatest lower bound. On the other hand, (29) shows that (30) is obtained by substituting into the function (5) of $(x_1, \cdots, x_n)$ the functions $x_i = x_i(t^*)$, functions which are continuous and periodic, hence bounded, for $-\infty < t^* < +\infty$. Consequently, if m denotes the greatest lower bound of the almost periodic function $r(t^*)$ for $-\infty < t^* < +\infty$, then $m \neq 0$, i. e., $m > 0$. Since $m \leqq \mu$ in view of (33_2), it follows that $\mu > 0$. Accordingly, one can assume without loss of generality that $\mu = 1$

(correspondingly, one could have modified the definition (7) of t^* by writing cr instead of r, where c is an arbitrarily positive constant). Thus

$$(34) \qquad 0 < m = \operatorname{fin\,inf} r(t^*) \leqq M\{r\} = \mu = 1.$$

Let $\phi(t^*)$ denote the function (32_2) in case of the normalization $\mu = 1$, so that, from (32_1),

$$(35) \qquad t \equiv t(t^*) = t^* + \phi(t^*).$$

According to (33_1), (32_2), the function $\phi(t^*) \equiv v(t^*)$ is almost periodic and has frequencies contained in the modul of the n (not necessarily linearly independent) numbers σ_i which are defined by (27_2).

It follows that one can define for $-\infty < t < +\infty$ a unique function $\psi(t)$ by the requirement (26), and that this $\psi(t)$ is almost periodic and has frequencies which are contained in the modul generated by the σ_i. In fact, $t'(t^*) = r(t^*)$ by (10), while $t'(t^*) = 1 + \phi'(t^*)$ by (35). Hence (34) can be written in the form

$$(36) \qquad 0 < m = 1 + \operatorname{fin\,inf} \phi'(t^*) \leqq 1.$$

Now there are two cases possible according as $m \neq 1$ or $m = 1$. If $m \neq 1$, then, from (36),

$$m - 1 = \operatorname{fin\,inf} \phi'(t^*), \text{ where } -1 < m - 1 < 0.$$

This means that (25) is satisfied by $\vartheta = 1 - m$, and so one can apply the general theorem concerning the inversion (26) of (35). If, on the other hand, $m = 1$, then (34) reduces to

$$1 = \operatorname{fin\,inf} r(t^*) = M\{r\},$$

so that $r(t^*) \equiv 1$ by the uniqueness theorem of almost periodic functions. Hence it is seen from (10) that the exceptional case $m = 1$ belongs to the trivial case where $\psi(t)$ in (26) is a constant.

Since, from (27_1) and (26),

$$u_i = \sigma_i t + \sigma_i \psi(t) + Q_i(t + \psi(t)), \text{ where } Q_i(t + \psi(t) + \tau_i) = Q_i(t + \psi(t)),$$

and since the frequencies of the almost periodic function $\psi(t)$ are contained in the modul generated by the $\sigma_i = 2\pi/\tau_i$, where $i = 1, \cdots, n$, it is clear from (23_1) that each of the n components $x_i(t)$ of the solution of the original Lagrangian system $[L]_{x_i} = 0$ is an almost periodic function whose frequencies are contained in the modul of the σ_i. In other words, the $x_i(t)$ possess an anharmonic Fourier analysis of the form

$$x_i(t) = \Theta_i(t/\tau_1, \cdots, t/\tau_n), \qquad (i = 1, \cdots, n), \tag{37}$$

where the

$$\Theta_i = \Theta_i(\theta_1, \cdots, \theta_n), \qquad (i = 1, \cdots, n), \tag{38}$$

are continuous functions of the position on the n-dimensional torus

$$0 \leqq \theta_1 < 1, \cdots, 0 \leqq \theta_n < 1,$$

while the $\tau_i = 2\pi/\sigma_i$ are the positive constants defined by (22). This is the desired result.

It is, in this connection, natural to ask whether or not the continuous function $\Theta = \Theta(\theta_1, \cdots, \theta_n)$ on the torus is necessarily regular analytic whenever it belongs to an almost periodic function $x(t)$ which is regular analytic and bounded in a strip about the t-axis and has n linearly independent frequencies. This problem seems to be quite difficult, since it is, in the main, a generalization of the (unsolved) problem of Poincaré-Denjoy in the analytic case ($n = 2$). The solution of the problem would be essential also for Levi-Civita's recent theory [6] of conditionally periodic systems, since, in general, nothing is known about $\Theta(\theta_1, \cdots, \theta_n)$ except for its continuity, so that not even Birkhoff's formal approach [1] to the Weierstrass preparation theorem is applicable.

As mentioned in the introduction, Stäckel [7] has found that a certain Jacobian cannot be distinct from zero for every t. A comparison of the procedure of the present paper with the calculations of Stäckel shows that the vanishing of the Jacobian in question is obvious without any particular calculations, since it is nothing but a manifestation of Hill's manifold of zero velocity. Correspondingly, the calculations and verifications of Stäckel [7] can be avoided by realizing that Hill's manifold of zero velocity cannot be reached by a path in a direction distinct from the transversal direction.†

THE JOHNS HOPKINS UNIVERSITY.

---

† This fact, well known in case of the restricted problem of three bodies (a case in which the transversal is the normal, the metric being Euclidean), holds in the general case (1) and can be proved as follows: Let the initial conditions x^0, x'^0 assigned to $t = 0$ be such that the vector x'^0 vanishes, so that x^0 is, by (2), a point of Hill's manifold $e(x) + h = 0$ of zero velocity belonging to $h = -e(x^0)$. For typographical reasons, differentiations with respect to t are denoted by primes instead of by dots. In order to exclude the case of an equilibrium solution $x(t) = \text{const.}$ represented by the

REFERENCES.

[1] G. D. Birkhoff, "Surface transformations and their dynamical applications," *Acta Mathematica,* vol. 43 (1922), pp. 1-119.

[2] G. D. Birkhoff, "Dynamical systems with two degrees of freedom," *Transactions of the American Mathematical Society,* vol. 18 (1917), pp. 199-300.

[3] H. Bohr, "Kleinere Beiträge zur Theorie der fastperiodischen Funktionen IV," *Det Kgl. Danske Videnskabernes Selskab. Mathematisk-fysiske Meddelelser,* vol. 10, no. 12 (1931), pp. 10-15.

[4] H. Bohr and B. Jessen, "Über fastperiodische Bewegungen auf einem Kreis," *Annali della R. Scuola Normale Superiore di Pisa,* ser. 2, vol. 1 (1932), pp. 385-398.

[5] J. Hadamard, "Sur les trajectoires de Liouville," *Bulletin des Sciences Mathématiques,* ser. 2, vol. 35 (1911), pp. 106-113.

[6] T. Levi-Civita, "Ergänzende Bemerkung zum Weierstrasschen Vorbereitungssatz und bedingt-periodische Bewegungen," *Zeitschrift für angewandte Mathematik und Mechanik,* vol. 13 (1933), pp. 112-114.

[7] P. Stäckel, "Über die geodätischen Linien einer Klasse von Flächen, deren Linienelement den Liouvilleschen Typus hat," *Crelle's Journal für Mathematik,* vol. 130 (1905), pp. 89-112.

---

single point $x = x^0$, one has to assume that grad $e(x)$ does not vanish at $x = x^0$. Then the hyper-surface $e(x) + h = 0$ has at $x = x^0$ an orientable normal and

$$\Sigma\Sigma\, g^{ik}(x^0)\, e_{x_i}(x^0)\, e_{x_k}(x^0) > 0.$$

Transversality being meant with reference to the Riemannian geometry of the $g_{ik}(x)$ which determine the quadratic part of (1), one has to prove that

$$\frac{|\Sigma\, x'_i(t)\, e_{x_i}(x^0)|}{\{\Sigma\Sigma\, g_{ik}(x^0)\, x'_i(t)\, x'_k(t)\}^{\frac{1}{2}}\, \{\Sigma\Sigma\, g^{ik}(x^0)\, e_{x_i}(x^0)\, e_{x_k}(x^0)\}^{\frac{1}{2}}} \to 1 \text{ as } t \to \pm 0,$$

where $x(t)$ is the solution for which $x(0) = x^0$, $x'(0) = x'^0 =$ zero vector. Since $\| g^{ik} \| = \| g_{ik} \|^{-1}$, it follows that it is sufficient to prove the relation

$$x'_i(t) = t \sum_k g^{ik}(x^0)\, e_{x_k}(x^0) + o(|t|), \text{ where } t \to \pm 0.$$

Now $x'_i(t) = x''_i(0)\, t + o(|t|)$, by Taylor's theorem. Hence it is sufficient to prove that

$$x''_i(0) = \sum_k g^{ik}(x^0)\, e_{x_k}(x^0), \text{ i.e., } e_{x_i}(x^0) = \sum_k g_{ik}(x^0)\, x''_k(0), \text{ where } i = 1, \cdots, n.$$

And the truth of the last relation follows by placing $t = 0$ (hence $x'^0_i = 0$) in the Lagrangian equations belonging to (1). In fact, these equations are easily found to be

$$\sum_k g_{ik} x''_k + \sum_j \sum_k \Gamma_i^{jk} x'_j x'_k + \sum_k \Pi_{ik} x'_k - e_{x_i} = 0, \qquad (i = 1, \cdots, n),$$

where the $\Gamma_i^{jk} = \Gamma_i^{kj}$ are the Christoffel symbols of the first kind belonging to the g_{ik}, while the $\Pi_{ik} = -\Pi_{ki}$ are the components of the alternating derivative (curl) of the covariant vector $(f_1, \cdots, f_n)$.

GALILEI GROUP AND LAW OF GRAVITATION.*

By Aurel Wintner.

It is known that among all attraction laws in which the force is proportional to a fixed, say the β-th, power of the distance, the case $\beta = -3$ is exceptional in many respects. For instance, in the case of two bodies all non-circular solutions lead, if $\beta = -3$, on the one hand to a collision and on the other hand to a recession to infinity. In the case of three bodies, Lagrange[1] has shown that in Newton's case $\beta = -2$ all homographic solutions are coplanar solutions; and this holds, according to Banachiewitz,[2] for every $\beta \neq -3$ but not for $\beta = -3$. The object of the following considerations is a general elucidation of the exceptional behavior of the inverse cubic law of attraction.

Since the ten classical integrals do not depend on the choice of the law of attraction and are due, according to Jacobi and Lie, to the infinitesimal trans-

---

* Received January 17, 1938.

[1] This is the main theorem of Lagrange concerning the homographic solutions of the problem of three bodies; cf., the concluding remarks on p. 292 of vol. 6 (1873) of Lagrange's collected works. A simple proof of the theorem is due to Pizzetti ("Casi particolari del problema dei tre corpi," *Rendiconti della Reale Accademia dei Lincei*, ser. 5, vol. 13_1 (1904), pp. 17-26), who also proved the more general theorem that every homographic solution of the problem of $n \geqq 2$ bodies is either coplanar or homothetic. This method of Pizzetti was rediscovered by Müntz (*Mathematische Zeitschrift*, vol. 15 (1922), pp. 169-187) and recently by Carathéodory (*Sitzungsberichte der Bayerischen Akademie der Wissenschaften*, 1933, pp. 257-267).

[2] T. Banachiewitz, "Sur un cas particulier du problème des trois corps," *Comptes Rendus*, vol. 142 (1906), pp. 510-512. The discussion of geometrical compatibility is not carried out by Banachiewitz but becomes quite easy by using the important fact, apparently not observed, that the non-planar solutions of Banachiewitz are isosceles (but not equilateral) solutions. For let $(a_i, b_i, 0)$ be the initial values of the barycentric coördinates (x_i, y_i, z_i) of the mass m_i, where $i = 1, 2, 3$, and let r_i denote the initial length $\{(a_j - a_k)^2 + (b_j - b_k)^2\}^{\frac{1}{2}}$ of a side of the triangle formed by the three bodies. Banachiewitz assumes that the three pairs (a_i, b_i) satisfy the conditions

$$\Sigma\, m_i a_i b_i = 0, \quad \Sigma\, m_i a_i b_i / r_i^4 = 0; \qquad \Sigma\, m_i a_i = 0, \quad \Sigma\, m_i b_i = 0,$$

and that the three points (a_i, b_i) of the initial plane are neither collinear nor such that $r_1 = r_2 = r_3$. Now, it is easily shown either by a determinant calculation or by an equivalent geometrical consideration that these conditions cannot be satisfied unless at least one of the $3 + 3$ numbers a_i, b_i vanishes, in which case direct substitution of $a_3 = 0$ into the above conditions shows that $r_1 = r_2$ (and $b_1 = b_2$, $m_1 = m_2$). Since the solution is homographic, it follows that the triangle is isosceles not only at $t = 0$ but for every t.

formations which generate the Galilei group,[3] it is clear that in the case of suitably chosen laws of attraction the group of transformations of the equations of motion is larger than the Galilei group. (If, for instance, $\beta = 2$, the problem of n bodies can be split into n vectorial harmonic oscillators and admits, therefore, certain continuous groups of orthogonal substitutions which it does not admit in case of another attraction law). Of course, this cannot be seen from the usual introduction of the Galilei group, since usually it is merely stated, but not proved, that the Galilei group exhausts the automorphic transformations of the Newtonian problem of n bodies. It will be seen that this implication concerning the completeness of the Galilei group, while (in the main) correct, is by no means so evident as to need no proof. In fact, it turns out that if the attraction were proportional to the —3rd, instead of the —2nd, power of the distance, the Galilei group would be only a subgroup of all "inertial" transformations involving the time.

Without assuming the existence of any of the ten classical integrals, consider a system of $k(=3n)$ differential equations of the second order, $x_i'' = f_i$, where $' = d/dt$ and $i = 1, \cdots, k$, and suppose that the k given continuous functions f_i are independent of t and are homogeneous, of some fixed degree β, in $(x_1, \cdots, x_k)$. For instance, $\beta = -2, 1, -1$ in the cases of Newton, Hooke and a logarithmical potential respectively. It is natural to seek pairs of *fixed* functions $u = u(t)$, $v = v(t)$ which have the property that $x_i = v(t)x_i(u(t))$ is, for *every* solution $x_i = x_i(t)$ of $x_i'' = f_i$, again a solution ("dynamical similarity"). It will be assumed that the solutions of $x_i'' = f_i$ are uniquely determined by the initial values and that $u(t)$, $v(t)$ have continuous second derivatives u'', v'', finally that v and the first derivative of u are positive on the t-interval under consideration. In particular, one can introduce $u = u(t)$ instead of t as an independent variable, so that $t = t(u)$.

Since the $f_i(x_1, \cdots, x_k)$ are homogeneous of degree β, it is found by direct substitution that if $x_i = x_i(t)$ is a fixed solution of $x_i'' = f_i$, where $x_i'' = d^2x_i/dt^2$, then $x_i = v(t)x_i(u(t))$ is again a solution if and only if

$$v^\beta \frac{d^2x_i}{du^2}\left(\frac{dt}{du}\right)^3 = \frac{d^2(vx_i)}{du^2}\ \frac{dt}{du} - \frac{d(vx_i)}{du}\ \frac{d^2t}{du^2}$$

is, in virtue of $t = t(u)$, an identity in u. It follows, therefore, by comparison of the coefficients of d^jx_i/du^j, where $j = 0, 1, 2$, that the two functions u, v

---

[3] Cf., F. Klein, *Vorlesungen über die Entwicklung der Mathematik im 19. Jahrhundert*, vol. II (1927), pp. 53-59.

of t will have the desired property with reference to every solution $x_i = x_i(t)$ of $x_i'' = f_i$ if the two functions t, v of u satisfy the three conditions

$$v/\dot{t}^2 = v^\beta, \qquad 2\dot{v}\ddot{t} - v\dddot{t} = 0, \qquad \ddot{v}\dot{t} - \ddot{t}\dot{v} = 0,$$

where the dots denote differentiations with respect to u. The third of these conditions means that $\dot{v}/\dot{t}$ is a constant, say c. Hence, on differentiating the first condition with respect to u and substituting the representation of $\ddot{t}$, thus obtained, into the second condition, it is seen that the three conditions can be written as follows:

$$\text{(I)} \quad \dot{t}^2 = v^{1-\beta}; \qquad \text{(II)} \quad 4\dot{t}^2\dot{v} = (1-\beta)v^{1-\beta}\dot{v}; \qquad \text{(III)} \quad \dot{v} = c\dot{t}.$$

Since $v > 0$ and $\dot{t} > 0$ by assumption, (I) implies that (II) is satisfied if and only if either $\beta = -3$ or $\dot{v} \equiv 0$.

Suppose first that $\dot{v} \equiv 0$ (for which $\beta \neq -3$ is sufficient but not necessary). Then (II) is an identity, while (III) is satisfied by $c = 0$ (and, since $\dot{t} > 0$, only by $c = 0$); so that the three conditions reduce to (I). Now $\dot{v} \equiv 0$, i. e., v is a (positive) constant, say λ. Hence, the single condition (I), where $\dot{t} = dt/du$, gives $u = \lambda^{\frac{1}{2}(\beta-1)}t$ plus an additive constant (which is unessential, since t does not occur explicitly in $x_i'' = f_i$). This determines the most general pair $u(t)$, $v(t)$ $(\equiv \lambda)$, if $\beta \neq -3$. Choosing, in particular, $\beta = -2$, it follows that $x_i = \lambda x_i(\lambda^{-3/2}t)$ is, for every positive constant λ and for every solution $x_i = x_i(t)$ of the problem of $n(=\frac{1}{3}k)$ bodies, a solution of the problem of n bodies.

This trivial transformation group, while not contained by the Galilei group, does not lead to an eleventh integral, a failure which, for reasons explained by Engel,[4] does not contradict Lie's integration theory.

Now let $\beta = -3$. Then (II) is, in virtue of (I), an identity also when $\dot{v} \not\equiv 0$, so that one can choose the constant c of (III) distinct from 0. Thus the three conditions (I), (II), (III) for the function pair $u = u(t)$, $v = v(t)$ reduce to the pair of conditions $u'^2 = v^{\beta-1}$, $v'\dot{t} = c\dot{t}$, or, since $\beta = -3$ and $\dot{t} > 0$, to $u' = v^{-2}$, $v' = c$; so that

$$u = u(t) = \int_0^t (c\bar{t} + b)^{-2}d\bar{t} + a, \qquad v = v(t) = ct + b,$$

where $a, b, c (\neq 0)$ are arbitrary constants. In particular, $x_i = -tx_i(t^{-1})$ is, for every solution $x_i = x_i(t)$ of $x_i'' = f_i$, again a solution,[5] and so the group of inertial transformations is essentially larger than the Galilei group.

---

[4] F. Engel, "Nochmals die allgemeinen Integrale der klassischen Mechanik," *Göttinger Nachrichten, 1917*, pp. 189-198.

This result seems to be of interest also because it explains *why* Jacobi was able to find, for the problem of n bodies in case of inverse cubic attraction, two new first integrals.[6] These allowed him to reduce the rectilinear problem belonging to $n = 3$ to a quadrature.[7]

THE JOHNS HOPKINS UNIVERSITY.

---

[5] This fact corresponds to the results of Cunningham and Bateman concerning the invariance of Maxwell's equations under certain transformations by reciprocal radii; cf. F. Klein, *loc. cit.*[3], pp. 78-79.

[6] Jacobi, *Gesammelte Werke*, Supplementband (1884), p. 27.

[7] Jacobi, *loc. cit.*[6], vol. 4 (1886), pp. 481-484 and pp. 533-539.

INTERIOR TRANSFORMATIONS ON SURFACES.*

By G. T. WHYBURN.

A single valued continuous transformation $T(A) = B$ is said to be *interior*[1] provided the image of every open set in A is open in B; and such a transformation is said to be *light*[2] provided the inverse set of every point in B is totally disconnected.

Stoilow[1] has analyzed light interior transformations under the assumption that both A and B are regions on a sphere or plane. Thus in this important case we have an analysis of interior light transformations from one (open) 2-dimensional manifold to another. In the present paper the principal theorem to be established (see § 2) is to the effect that if A is a compact 2-dimensional manifold (with or without boundary curves), so also is any light interior image of A. Thus the 2-dimensional manifold character of a set appears as an invariant under light interior transformations and accordingly it need not be assumed for the image set.

Various applications of our principal theorem will follow in §§ 3 and 4. Notably, in case A is a sphere, it is shown that any light interior image of A is necessarily either a sphere, projective plane or 2-cell; and furthermore each true cyclic element of *any interior image* of A (whether the transformation is light or not) is a sphere, projective plane or 2-cell.

1. Preliminary lemmas and theorems.

All sets considered are assumed to lie in a separable metric space. We begin with

(1.1). LEMMA. *If M is a locally connected locally compact continuum having no local separating point, each point $p \in M$ is contained within an arc qpr of M which does not separate M.*

Proof. There exists[3] an uncountable aggregate X of arcs $[px]$ in M each pair of which intersect in just p. Since the aggregate of disjoint connected

---

* Received January 13, 1938.

[1] See Stoilow, *Annales Scientifiques de l'École Normale Supérieure*, vol. 63 (1928), pp. 347-382 and *Annales de l'Institut Henri Poincaré*, vol. 2 (1932), pp. 233-266.

[2] See my paper in the *Duke Mathematical Journal*, vol. 3 (1937), pp. 370-381.

[3] See G. T. Whyburn, *American Journal of Mathematics*, vol. 53 (1931), pp. 305-314; see also an abstract by Zippin in the *Bulletin of the American Mathematical Society*, vol. 36 (1930), p. 805.

subsets $[px - p]$ of the connected set $M - p$ is uncountable it therefore contains[4] an uncountable subaggregate Y such that for each set $py - p$, in Y, $(M - p) - (py - p)$ has at most two components and every such component is bounded by the entire set $py - p$. On any one such arc py choose an interior point q. Then clearly $M - pq = (M - p) - (pq - p)$ is connected. Let Z denote the collection obtained by omitting from Y the set containing q. Then since Z is an uncountable collection of disjoint connected subsets of the connected set $M - pq$ it therefore contains at least one element, say $pz - p$ such that $(M - pq) - (pz - p)$ has at most two components and each of these is bounded (rel. $M - pq$) by all of $pz - p$. Let r be any interior point of pz. Then clearly

$$(M - pq) - (pr - p) = M - (pq + pr) = M - qpr$$

is connected.

(1.2). *Let $T(A) = B$ be interior and light where A is connected and locally connected and is locally a cantorian manifold*[5] *of dimension ≥ 2. Then B is also locally a cantorian manifold of dimension ≥ 2.*

Proof. For suppose some connected open subset Q of B is separated by a compact 0-dimensional subset X'. Then X' contains a closed subset X which irreducibly separates Q between some two points a and b. Let $y \in X$, $x \in T^{-1}(y)$ and let V be a connected neighborhood of x so chosen that $T(V) \subset Q$ and $F(V) \cdot T^{-1}(X) = 0$.[6] (Note $T^{-1}(X)$ is of dimension 0 since T is continuous and light.) Let $S = V - V \cdot T^{-1}(X)$. Then S is connected and open and hence so also is its image. But $T(S) \subset Q - X$ and since $T(S)$ must intersect both the component of $Q - X$ containing a and the one containing b it follows that X does not separate a and b in Q. Thus we have a contradiction and our result is proven.

(1.21). COROLLARY. *Let $T(A) = B$ be continuous and light where A is locally connected and is locally a cantorian manifold of dimension ≥ 2. If T is interior at the point $x \in A$, then $T(x)$ is not a local separating point of B. Thus in particular, if T is interior on A, B has no local separating points.*

---

[4] See my paper in the *Transactions of the American Mathematical Society*, vol. 33 (1931), pp. 444-454.

[5] A connected and locally connected set A is said to be locally a cantorian manifold of dimension ≥ 2 provided no connected open subset of A is separated by a compact, totally disconnected set.

[6] For any open set V, $F(V)$ denotes the set-theoretic boundary of V, i. e., the set $\bar{V} - V$. For any set X and any number $\delta > 0$, $V_\delta(X)$ denotes the set of all points at a distance $< \delta$ from X.

(1.3). THEOREM. *Let $T(A) = B$ be interior and light where A is a 2-dimensional manifold (with or without boundary curves). Then for each $b \in B$, $T^{-1}(b)$ is finite.*

Proof. By (1.21) B has no local separating point. Hence by (1.1) each point $b \in B$ is interior to some arc pbq in B such that $B - pbq$ is connected. Let $X = T^{-1}(pbq)$. Then $A - X$ can have only a finite number of components[2] and furthermore X is a locally connected set.[2] Hence X can contain at most a finite number of simple closed curves. Thus if C is any component of X, each true cyclic element of C is a graph and there are only a finite number of such cyclic elements. Accordingly $T^{-1}(b)$ can intersect each cyclic element of C in just a finite number of points and thus can contain only a finite number of points of C which are on true cyclic elements of C. But since b is of order 2 in pbq it follows that $C \cdot T^{-1}(b)$ contains no end points of C; and since $pbq - b$ has just two components, $T^{-1}(b) \cdot C$ can contain only a finite number of cut points of C. Therefore $T^{-1}(b) \cdot C$ is finite and hence so also is $T^{-1}(b) \cdot X = T^{-1}(b)$, since X has only a finite number of components C.

(1.4). *If A is a 2-cell with boundary J, if $T(A) = B$ is interior and light and if C is a simple closed curve in B such that $T^{-1}(C) = J$, then B is a 2-cell with boundary C.*

Proof. By virtue of Zippin's characterization of the 2-cell[7] we have to prove that every arc spanning C in B irreducibly separates C.

To that end let cvd be any such arc and let x and y be points separating c and d on C. We first show that cvd separates x and y in B. If this is not so we may suppose we have an arc $xy \subset B - cvd$ so that $xy \cdot J = x + y$. Let c_1 and d_1 be points of $T^{-1}(c)$ and $T^{-1}(d)$ respectively. An arc c_1d_1 of J contains at least one point of $T^{-1}(x)$ or $T^{-1}(y)$, say a point x_1 of $T^{-1}(x)$. There exists an arc x_1y_1 in $T^{-1}(xy)$ which maps topologically onto xy.[2] Since there exists similarly an arc d_1c_2 in $T^{-1}(cvd)$ with $c_2 \subset T^{-1}(c)$, it follows that the arc $x_1d_1y_1$ of J contains c_2. The subarc c_2d_1 of $x_1d_1y_1$ must contain a point of $T^{-1}(x)$ or $T^{-1}(y)$, say a point x_2 of $T^{-1}(x)$. Then, repeating the argument, it follows that c_2d_1 also contains a point y_2 of $T^{-1}(y)$; and then that the subarc x_2y_2 of c_2d_1 contains points c_3 and d_2 of $T^{-1}(c)$ and $T^{-1}(d)$ respectively and so on. Clearly this is impossible, since each of the sets $T^{-1}(c)$, $T^{-1}(d)$, $T^{-1}(x)$ and $T^{-1}(y)$ is finite. Hence cvd separates x and y in B. Clearly no subarc of cvd can separate x and y in B. Thus since B is unicoherent,[8] no proper subset of cvd separates x and y in B; and since clearly there cannot exist three finite sets of components of $A - T^{-1}(cvd)$ each having a subcontinuum of $T^{-1}(cvd)$ on its boundary, it follows that cvd separates B into just two components and hence separates it irreducibly. Accordingly B is a 2-cell with boundary C.

---

[7] *American Journal of Mathematics*, vol. 55 (1933), pp. 201-217.

[8] See Eilenberg, *Fundamenta Mathematicae*, vol. 24 (1935), p. 175

We conclude this section with a lemma which, while not used in the proof of our main result, will be useful in some of the applications.

(1.5). LEMMA. *Let $T(A) = B$ be interior, let R be an open subset of A with boundary F, let $B_0 = T(\bar{R})$ and designate the transformation $T(\bar{R}) = B_0$ by T^*. Let $E \subset F$ be the set of points (if any) where T^* fails to be interior. Then for any $x \in E$ and any neighborhood U of x, $T(E \cdot U)$ locally separates B_0 at the point $T(x)$.*

Proof. Let $y = T(x)$, let $\epsilon > 0$, let U be any open subset of A containing x such that $\bar{U} \subset V_\epsilon(x)$ and let $U_0 = U \cdot R$. Finally let $V = T(U_0)$. Now if we suppose $T(E \cdot U)$ does not locally separate B_0 at y, there will exist an open subset W of B_0 containing y and such that $W \subset T(U)$ and $W - W \cdot T(E) = X$ is connected. Then we must have $X \subset V$. For if not there will exist a point $z \in X \cdot V(\overline{X - X \cdot V}) + \overline{X \cdot V}(X - X \cdot V)$; and if $p \in U \cdot T^{-1}(z)$ is chosen so that $p \in \bar{U}_0$, then since $U_0 = U \cdot \bar{R}$ we must have $p \in U_0$ and hence $z \in X \cdot V(\overline{X - X \cdot V})$. But since $z \in X$ and $X \cdot T(E) = 0$, it follows that T^* is interior at p. Hence z must be interior to $T^*(U_0) = V$, contrary to $z \in \overline{X - X \cdot V}$. Thus $X \subset V$. Now since every point of $W \cdot T(E)$ is a limit point of X (and thus also of V) it follows that

$$\bar{V} = T(\bar{U}_0) \supset W.$$

Whence

$$T[V_\epsilon(x) \cdot \bar{R}] \supset T(\bar{U}_0) \supset W.$$

Thus $T(x)$ is interior to the image of arbitrarily small open subsets of $\bar{R}$ containing x and accordingly $T^*(\bar{R}) = B_0$ is interior at x, contrary to $x \in E$.

(1.51). COROLLARY. *If a point x of E is an isolated point of E, $T(x)$ is a local separating point of B_0.*

2. THEOREM. *If A is a 2-dimensional manifold (with or without boundary curves) and $T(A) = B$ is interior and light, then B is a 2-dimensional manifold.*

For the purpose of this proof a point x will be called a *regular point* of A or B provided there exists a neighborhood U of x (in A or B) such that $\bar{U}$ is a 2-cell whose boundary curve does not contain x. Any non-regular point will be called *singular*.

The proof will be given in the form of a series of statements.

(i) *The boundary of any component of the complement of a compact locally connected set in B is locally connected.*

To prove this, let R be a component of $B - M$ where M is a compact locally connected set in B and let $F = F(R)$. Let $T^{-1}(M) = N$, let S be a component of $T^{-1}(R)$ and let $E = F(S)$. Then[2] N is locally connected. Also S is a component of $A - N$. Accordingly E is locally connected. Now since[2] $T(S) = R$ it follows that $T(E) = F$. Therefore F is locally connected, because T is continuous.

(ii) *For any $b \in B$ and any $\epsilon > 0$ there exists a simple closed curve or an arc X in B which irreducibly separates B into just two components and so that the component of $B - X$ containing b is of diameter $< \epsilon$.*

Proof of (*ii*). Let $a \in B - b$, let $b' \in T^{-1}(b)$ and let U be a 2-cell neighborhood of b' with boundary curve J such that the topological boundary $F(U)$ of U is either J or an arc cvd of J and such that $\bar{U} \cdot T^{-1}(b) = b'$, $\bar{U} \cdot T^{-1}(a) = 0$. [Note $T^{-1}(b)$ is finite by (1.2).] Now since, by (1.21), B has no local separating point, there exists a locally connected subcontinuum M of B which ϵ-separates b in B and is such that $T^{-1}(M) \cdot U \neq 0$, $T^{-1}(M) \cdot F(U) = 0$, and such that the component S^* of $B - M$ containing a also contains $T[F(U)]$. Let R be the component of $B - \bar{S}^*$ containing b, let S be the component of $B - \bar{R}$ containing b and let X be the common boundary of R and S. Then X is a compact locally connected subset of M [by (i)] which irreducibly separates a and b in B and $\delta(R) < \epsilon$. Now if R' and S' denote the components of $T^{-1}(R)$ and $T^{-1}(S)$ containing b' and $F(U)$ respectively and if $Z = U \cdot T^{-1}(X)$, it follows that since $b' = U \cdot T^{-1}(b)$ and $T^{-1}(a) \cdot \bar{U} = 0$ there can be no other components of either $T^{-1}(R)$ or $T^{-1}(S)$ intersecting $\bar{U}$. Accordingly $Z = F(R')$ and $Z \subset F(S')$. Hence Z is a continuum and since it is locally connected (because X is locally connected) it must be either a simple closed curve or a simple arc according as b' is a regular point or a singular point of A. Furthermore, since $Z = F(R')$ and $T(R') = R$ it follows that $T(Z) = X$, and since the transformation $T(Z) = X$ is interior on Z, X is either an arc or a simple closed curve.[2] Finally, since there can exist no component of $A - T^{-1}(X)$ other than R' and S' with boundary points in Z, clearly $R + S = B - X$ so that X irreducibly separates B.

Now continuing with the same notation we next prove

(iii) *If X is a simple closed curve, $R + X = \bar{R}$ is a 2-cell with boundary curve X; and thus b is a regular point.*

For if X is a simple closed curve, Z must be a simple closed curve and $R' + Z = \bar{R}'$ is a 2-cell with boundary curve Z. Furthermore the transformation $T(\bar{R}') = \bar{R}$ is interior, since $\bar{R}' = U \cdot T^{-1}(\bar{R})$. Thus since $Z = \bar{R} \cdot T^{-1}(X)$ it follows by (1. 4) that $\bar{R}$ is a 2-cell with boundary curve X.

(iii′) *Every singular point of A maps into a singular point of B.*

This is a corollary to (iii).

Still retaining the notation of (ii) we next establish

(iv) *A point x of X is a singular point of B if and only if it is an end point of X.*

First let x be a non-end-point of X and let $q \epsilon T^{-1}(x)$. Let σ be arbitrarily small. Since x is not a cut point of B there exists a subcontinuum N of B such that $B - x \supset N \supset B - V_{\sigma/2}(x)$ and $N \supset T(u) + T(v)$, where u and v are points of z so chosen that the arc uqv of z maps topologically[2] into an arc exf of X under T. Now we can construct in A an arbitrarily small simple closed curve $C = rhs + rks$ such that rhs and rks are arcs in $R' + r + s$ and $S' + r + s$ respectively, $r + s = C \cdot Z$, and $T(C) \cdot N = 0$. Now since C separates x and u in A it follows that $T(C)$ must separate $T(x)$ and N in B. Hence by the argument used in proving (ii), $T(C)$ must contain a set X' irreducibly separating x and N in B and X' is either an arc or a simple closed curve. But now since the set $T^{-1}(X')$ must separate x and u in A and T is topological on uqv, it follows that $T^{-1}(X')$ contains continua H and K such that $r + s \subset H \subset R' + r + s$, $r + s \subset K \subset S' + r + s$. Accordingly X' contains the continua $T(H)$ and $T(K)$. Also $T(H) \cdot T(K) = T(r) + T(s)$ and $T(r) \neq T(s)$. Hence X' cannot be an arc and thus it is a simple closed curve. Therefore, by (iii), x is a regular point.

Now if x is an end point of X, since clearly x can have no 2-cell neighborhood not having x on its boundary curve it follows that x is a singular point.

Now let K denote the set of all singular points of B. Clearly K is closed. It may be vacuous in case A has no singular points, but not otherwise, by (iii′).

(v) *Every point of K is of order 2 (rel. K).*

Proof. By (iv) it follows that every point of K is of order ≤ 2. Now if there were a point x of K of order < 2, there would exist an arbitrarily small locally connected continuum X in B such that $X \cdot K$ contains < 2

points and X irreducibly ϵ-separates x in B. By (iii) it follows that X is an arc and by (iv) that both its end points must belong to K, contrary to supposition.

(v′) *Each component of K is a simple closed curve.*

(vi) *$B - K$ is connected.*

If this were not so we could find a subset K' of K and two distinct components R and S of $B - K'$ with a boundary point p in common, $p \epsilon K'$. But by (ii) and (iv) there exists an arbitrarily small arc ab in B with $ab \cdot K = a + b$ and so that ab separates p from both r and s where $r \epsilon R$, $s \epsilon S$. Clearly this is impossible, since we would then have $ab - (a + b) \subset R \cdot S$.

(vii) *K has only a finite number of components. In fact K has at most p components,*[9] *where $p - 1$ is the maximum number of disjoint simple closed curves in A whose sum does not separate A.*

To prove this, we note first by (vi) it follows that there are only a finite number of components of $A - T^{-1}(K)$. Thus since each component of $T^{-1}(K)$ contains[2] a simple closed curve and since A is disconnected by the removal of any set of p disjoint simple closed curves it follows that the number of components of $T^{-1}(K)$—and hence also of K—must be finite.

To show that the number of components of K does not exceed p, let us suppose the contrary. Let F_1 be the sum of a collection of p components of $T^{-1}(K)$ and let R_1 and S_1 be components of $A - F_1$. If R_1 contains a component of $T^{-1}(K)$, let F_2 be the sum of a set of p-components of $T^{-1}(K)$ selected so that $F_2 \subset \bar{R}_1$ and F_2 conatins at least one component of $T^{-1}(K)$ lying in R_1. Let R_2 and S_2 be components of $A - F_2$ where $S_2 \supset S_1$. Then $R_2 \subset R_1$. Similarly if R_2 contains a component of $T^{-1}(K)$ we select a set F_3 of p components of $T^{-1}(K)$ so that $F_3 \subset \bar{R}_2$ and F_3 contains at least one component of $T^{-1}(K)$ lying in R_2 and take components R_3 and S_3 of $A - F_3$ so that $S_3 \supset S_2$ and hence $R_3 \subset R_2$. Continuing this process, since the number of components of $T^{-1}(K)$ is finite, we eventually find a component R_k of $A - F_k$ such that $R_k \cdot T^{-1}(K) = 0$. But then R_k is a component of $A - T^{-1}(K)$. Hence[2] $T(R_k) = B - K$ and thus $T(F_k) = K$. Clearly this is impossible if K contains more than p components, since F_k has exactly p components.

(viii) *Every singular point of B has a 2-cell neighborhood.*

Proof. Let $p \epsilon K_i \subset K$ where K is the set of all singular points of B and

---

[9] Actually, $p \leqq p_2{}^1(A) + a + 1$, where a is the number of boundary curves of A and $p_2{}^1(A)$ is the Betti number, modulo 2, of A.

K_i is a component of K. Let u and v be points of $K_i - p$, let q be a point of $T^{-1}(p)$. Since $T^{-1}(K)$ is a finite graph, there exists a 2-cell neighborhood U of q with boundary J (which may or may not contain q) such that $\bar{U} \cdot T^{-1}(K)$ consists of a finite number of arcs $qx_0, qx_1, \cdots, qx_n$ such that (a) $qx_i \cdot qx_j = q$, (b) $x_0, x_1, \cdots, x_n$ are cyclicly ordered on J and the arc x_0x_n of J not containing q is the set theoretic boundary of U in case q is on J, (c) qx_0 maps topologically into a subarc either of pu or pv, say pu, qx_1 maps topologically onto a subarc of pv, qx_2 into a subarc of pu, and so on to qx_n, where pu and pv are the arcs of K_i not containing v and u respectively, (d) $\bar{U} \cdot T^{-1}(p) = q$.

Now let ab be an arc in B irreducibly ϵ-separating p in B where $a \epsilon pu$, $b \epsilon pv$ and where ϵ is so chosen that the component R of $B - ab$ containing p satisfies $\bar{R} \cdot T(\bar{U} - U) = 0$. Let S be the other component of $B - ab$. Since $T^{-1}(S) \supset x_0x_1$ (arc of J) and $T^{-1}(R) \supset q$, it follows that the open 2-cell V in U bounded by the simple closed curve $G = qx_0 + x_0x_1 + qx_1$ contains a component R' of the inverse of the connected region $Q = R - \widehat{apb}$ (open arc of K_i). Since $x_0x_1 \subset T^{-1}(S)$, it follows at once that the boundary of R' is a simple closed curve $W = a'q + qb' + a'b'$, where $a'q$ and qb' are arcs on qx_0 and qx_1 respectively and where $a'b'$ is an arc which maps into ab and lies except for a' and b' in V. Since $a'b'$ is the common boundary of R' and the component S' of $T^{-1}(S)$ containing x_0x_1 and $a' = T^{-1}(ab) \cdot qx_0$, $b' = T^{-1}(ab) \cdot qx_1$, it follows that $a'b' = T^{-1}(ab) \cdot \bar{V}$ and hence that $a'b'$ maps interiorly onto ab. Thus since $T^{-1}(a) \cdot a'b' = a'$, $T^{-1}(b) \cdot a'b' = b'$ it follows that $a'b'$ maps topologically onto ab.

Therefore we have shown that the boundary W of the open 2-cell R' maps topologically into the boundary C of the region Q while R' maps onto Q. Hence surely the transformation $T(\bar{R}') = \bar{Q}$ is interior, and since $T^{-1}(C) \cdot \bar{Q} = W$, it follows by (1.4) that $\bar{Q}$ is a 2-cell with boundary C.

Clearly the statements (i)–(viii) yield our theorem.

3. Theorem. *If A is a sphere and $T(A) = B$ is interior and light then B is either a 2-cell, a sphere or a projective plane*[11] *(a 2-cell if B has singular points, otherwise a sphere or projective plane).*

Proof. If B has no singular point then by § 2 B is a closed 2-dimensional manifold without boundary; and since by a theorem of the author's[10] $p^1(B)$

---

[10] See my paper "On the mapping of Betti groups under interior transformations," *Duke Mathematical Journal*, vol. 4 (March, 1938). Here $p^1(X)$ denotes the Betti number (modulo 0) of a set X.

[11] That each of these image sets is possible is shown incidentally in the proof of (3.1).

$\leq p^1(A) = 0$, it follows that B is either a sphere or a projective plane according as it is or is not orientable.

If B has singular points, then by § 2, (vii), the set K of such singular points is a simple closed curve. Let R be a component of $A - T^{-1}(K)$ and let C be the boundary of $\bar{R}$, where R is so chosen that $A - \bar{R}$ is connected. (There are only a finite number of components of $A - T^{-1}(K)$ since by § 2, (vi), $B - K$ is connected.) Then since[2] $T^{-1}(K)$ is locally connected it follows that C is a simple closed curve and $\bar{R}$ is a 2-cell. Furthermore $T(\bar{R}) = B$ and $T^{-1}(K) \cdot \bar{R} = C$; and since no subset of $T(C) = K$ can locally separate B at any one of its points it follows by (1. 5) that the transformation $T(\bar{R}) = B$ is interior. Therefore, by (1. 4), B is a 2-cell.

(3. 1). Corollary. *Let $T(A) = B$ be interior and light. If A is a 2-cell, so also is B; if A is a sphere or projective plane, B is either a 2-cell, sphere or projective plane.*

Proof. First let A be the 2-cell $x^2 + y^2 \leq r^2$ in the (x, y) plane. Then the transformation $T'(x, y, z) = (x, y, 0)$ on the sphere A': $x^2 + y^2 + z^2 = r^2$ is interior and light and maps A' into A. Hence TT' is interior and light and maps A' into B; and since by § 2, (iii′), B necessarily has singular points, it follows from the above theorem that B is a 2-cell.

Next let A be a projective plane. The transformation T' obtained on a sphere A' by identifying diametrically opposite points is interior and light and maps A' into a projective plane which we may suppose is A. Then TT' is interior and light and maps A' onto B and hence again our result follows from the theorem just proved.

Finally, the case of the sphere is identical with the above theorem.

4. Applications to other surfaces.

(4. 1). Theorem. *Let $T(M) = N$ be continuous and light where M is a locally connected continuum and suppose that for each $y \in N$, $T^{-1}(y)$ disconnects no cyclic element of M. Then for each A-set*[12] *N_a in N, each component of $T^{-1}(N_a)$ is an A-set in M.*

Proof. Let K be a component of $T^{-1}(N_a)$. Let R be any component of $M - K$. Now $F(R)$ must reduce to a single point. For if not, then since any two points of $F(R)$ are conjugate,[12] there exists a cyclic element E of M containing $F(R)$. Since $R \cdot E$ is non-vacuous and connected, it follows that $R \cdot E$ is not a subset of $T^{-1}(N_a)$. Let S be a component of $E - E \cdot T^{-1}(N_a)$. Then $T(S)$ is a connected subset of $N - N_a$ and hence $F(S)$ must map into the single point p which is the boundary of the component of $N - N_a$ containing $T(S)$. Therefore $F(S)$ is totally disconnected and is $\subset T^{-1}(p)$; and since $K \cdot E$ is connected and non-degenerate, $T^{-1}(p)$ must disconnect E, contrary to hypothesis. Hence $F(R)$ is a single point and thus[12] K is an A-set.

---

[12] An A-set in a locally connected continuum M is a closed subset K of M which contains every arc in M whose endpoints lie in K. See Kuratowski and Whyburn, *Fundamenta Mathematicae*, Vol. 16 (1930), pp. 305-331.

(4.11). Corollary. *If $T(M) = N$ is continuous and light where M is a locally connected continuum such that*

(a) *M is unicoherent or*

(b) *every true cyclic element is a cantorian manifold of dimension ≥ 2 then each component of the inverse of an A-set in N is an A-set in M.*

(4.12). Corollary. *If $T(M) = N$ is interior and light and for no $y \in N$ does $T^{-1}(y)$ disconnect any cyclic element in M, then for each A-set N_a in N, $T^{-1}(N_a)$ is the sum of a finite number of disjoint A-sets in M, each of which maps interiorly onto N_a.*

(4.2). Theorem. *If $T(A) = B$ is interior and light where A is a locally connected continuum which either*

(a) *is unicoherent, or*

(b) *has as its true cyclic elements sets which locally are cantorian manifolds of dimension ≥ 2,*

then for each true cyclic element E_b in B there exists a true cyclic element E_a in A which maps interiorly onto E_b under T.

Proof. By the preceding theorem and corollaries, $T^{-1}(E_b)$ is made up of a finite number of disjoint A-sets each mapping onto E_b. Let E_a be a node[13] of any one of these components K of $T^{-1}(E_b)$. Then since the transformation $T(K) = E_b$ is interior it follows[10] that E_a maps onto E_b under T. Furthermore, the transformation $T(E_a) = E_b$ is interior except possibly at the one point $p = E_a \cdot \overline{K - E_a}$. If this transformation failed to be interior at p, there would exist a point $q \in E_a - p$ such that $T(q) = T(p)$. But then since $T(E_a) = E_b$ is interior at q it follows by (1.21) that $T(p)$ is not a local separating point of E_b; and hence by (1.51), $T(E_a) = E_b$ is interior at p. Thus $T(E_a) = E_b$ is interior and our theorem is proven.

---

[13] A node of a locally connected continuum M is either an end point of M or a true cyclic element of M containing exactly one cut point of M. See my paper in the *American Journal of Mathematics*, vol. 50 (1928), pp. 167-194.

This theorem, together with the results established in §§ 2, 3, yield the following corollaries.

(4. 21). *Under the conditions of the theorem, B likewise satisfies* (a) *or* (b) *respectively.*

(4. 22). *If every true cyclic element of a locally connected continuum A is a 2-dimensional manifold (with or without boundary curves) the same is true of any light interior image of A.*

(4. 23). *If every true cyclic element of a locally connected continuum is either a sphere, 2-cell or projective plane, the same is true of any light interior image of A.*

Now if we make use of the fact[2] that any interior transformation T can be factored into the form $T = T_2T_1$ where T_1 is monotone and T_2 is interior and light we get:

(4. 24). *If every true cyclic element of a locally connected continuum A is either a sphere or a 2-cell and if $T(A) = B$ is any interior transformation (light or not), then each true cyclic element of B is either a sphere, 2-cell, or projective plane.*

For if we factor T into T_2T_1 where $T_1(A) = A'$ is monotone and $T_2(A') = B$ is interior and light, it follows[14] that every true cyclic element of A' is either a sphere or a 2-cell. Hence by (4. 23) we get our conclusion.

(4. 3). THEOREM. *Let $T(A) = B$ be interior and light where A is a locally connected continuum and every true cyclic element of A is locally a cantorian manifold of dimension $\geqq 2$. Then for each true cyclic element E_b in B we have*

$$T^{-1}(E_b) = \sum_{i=1}^{k} E_a{}^i$$

where for each i, $E_a{}^i$ is a true cyclic element of A and the transformation $T(E_a{}^i) = E_b$ is interior.

Proof. Let K be any component of $T^{-1}(E_b)$. By (4. 1), K is an A-set in A. Since[10] each node of K maps onto all of E_b it follows that there are only a finite number of nodes of K. Hence if E_a is any true cyclic element in

---

[14] See R. L. Moore, *Transactions of the American Mathematical Society*, vol. 27 (1925), pp. 416-428; C. B. Morrey, *American Journal of Mathematics*, vol. 57 (1935), pp. 17-50; G. T. Whyburn, *ibid.*, vol. 56 (1934), pp. 294-302.

K, then E_a contains just a finite set F of cut points of K, i.e., the set $F = E_a \cdot \overline{K - E_a}$ is finite. Hence $T(F) = H$ is finite. Now, since, by (4.21), E_b is a cantorian manifold of dimension ≥ 2, it follows that $E_b - T(F)$ is connected; and since by hypothesis E_a is such a manifold, $E_a - E_a T^{-1}(H) = G$ is likewise connected. But G is a component of $K - K \cdot T^{-1}(H)$, since $T^{-1}(H) \supset F$. Hence[2] G maps onto all of $E_b - T(F)$ and thus E_a maps onto all of E_b under T. Since T is interior on $E_a - F$ and F is finite, and since E_b has no local separating point, it follows by (1.5) that T is interior on E_a.

Since each true cyclic element of K maps interiorly onto all of E_b under T, it follows that there are only a finite number of such elements in K. And since clearly K can have no arc of cut points [for $T(K) = E_b$ is interior], it follows that K is the sum of a finite number of true cyclic elements each mapping interiorly onto E_b. Thus since $T^{-1}(E_b)$ has only a finite number of components, our theorem follows.

(4.31). Corollary. *If every true cyclic element of A is a 2-dimensional manifold and $T(A) = B$ is interior and light, for each non-end-point $b \in B$, $T^{-1}(b)$ is a finite set.*

If b belongs to a true cyclic element of B, this follows from (4.3) and (1.2). If b is a cut point of B it results at once from the easily established fact that $T^{-1}(b)$ is contained in the sum of a finite number of cyclic elements of A.

(4.32). Corollary. *If every true cyclic element of A is a sphere or 2-cell and $T(A) = B$ is interior, then for each non-end-point $b \in B$, $T^{-1}(b)$ has only a finite number of components.*

This follows from (4.31) using the factorization of T into a monotone transformation and a light interior transformation just as was done in the proof of (4.24).

It may be remarked that it follows from the above results that any one dimensional interior image of a sphere—or of any locally connected continuum every true cyclic element of which is a sphere, 2-cell, or projective plane—is necessarily a dendrite. Also it is interesting to note that—by sending certain indecomposable continua into points—a sphere may be mapped interiorly into *any dendrite.* However, if we specify that our interior transformation shall send only locally connected sets into single points, it results that the only possible one-dimensional images of a 2-dimensional manifold are the *arc* and the *simple closed curve.* A detailed study of these and closely related results will be made in a later paper.

5. In conclusion, we shall isolate—essentially from the properties developed in § 2—a theorem which yields a complete analysis "in the small" of any light interior transformation on a 2-dimensional manifold. We begin by establishing

(5.1). THEOREM. *Let $T(A) = B$ be interior and light. Suppose E is a 2-cell with interior R and generator J in A such that $E \cdot T^{-1}T(J) = J$ and T is $(1-1)$ on J. Then T is $(1-1)$ on E, so that the transformation $T(E) = F$ is topological.*

Proof. Suppose, on the contrary, that there exist two points x and y in R such that $T(x) = T(y) = p$. Let apb be an arc in F such that $apb \cdot T(J) = a + b$. There exist arcs $a'xb'$ and $a'yb'$ in E which map topologically onto apb under T, where $a' = E \cdot T^{-1}(a)$, $b' = E \cdot T^{-1}(b)$. Since these two arcs are different, there exists a component S of $E - (a'xb' + a'yb')$ such that $S \cdot J = 0$ and hence so that $S \subset R$. Clearly this is impossible, since S contains a component U of $A - T^{-1}(apb)$ and U necessarily maps onto one of the components of $F - apb$ and each of these contains points of $T(J)$.

(5.2). THEOREM. *Let $T(A) = B$ be interior and light, where A is a 2-dimensional manifold. For any point q of A there exists a closed 2-cell neighborhood E of q in A and a positive integer k such that* (a) *if $T(q)$ is a regular point of B, then on E T is topologically equivalent to the transformation $w = z^k$ on $|z| \leqq 1$;* (b) *if $T(q)$ is a singular point of B, then on E T is topologically equivalent to the transformation*

$$f(z) = \rho(\cos k\theta/2 + i\,|\sin k\theta/2\,|) \qquad (k \text{ even})$$

on $|z| \leqq 1$ when q is a regular point of A and to this same transformation (for a different value of k) on $|z| \leqq 1$, $y \geqq 0$ when q is a singular point.

Proof. Let us consider first the case where $T(q)$ is a singular point of B. Then, referring back to the proof of (viii) in § 2, it follows by (5.1) that the set $\bar{R}'$ there defined maps topologically under T onto the set $\bar{Q}$. In that proof let us set $V = V_0$, $a' = a_0$, $b' = a_1$, $R' = R_0$. Let V_i be the open 2-cell in U bounded by the simple closed curve $qx_i + x_ix_{i+1} + qx_{i+1}$ for $i < n$ if q is on J and for $i \leqq n$ $(n + 1 \equiv 0)$ if q is within J. Then in exactly the same way it follows that there exist arcs $a_1a_2, a_2a_3, \cdots, a_ia_{i+1}, \cdots$ in $\bar{V}_1, \bar{V}_2, \cdots, \bar{V}_i, \cdots$ so that if R_i is the open 2-cell in V_i bounded by the simple closed curve $qa_i + a_ia_{i+1} + qa_{i+1}$ (where $0 \leqq i \leqq n - 1$ for q on J and $0 \leqq i \leqq n$, $n + 1 \equiv 0$, if Q is within J), $\bar{R}_i$ maps topologically onto $\bar{Q}$ under T. Hence if we call E the closed 2-cell $\Sigma\bar{R}_i$, it follows that, on E, T is topologically equivalent to the transformation $f(z) = \rho(\cos n\theta + i\,|\sin n\theta\,|)$ on $|\,z\,| \leq 1$, $y \geq 0$, if q is on J and to the transformation

$$f(z) = \rho[\cos(n+1)\theta/2 + i\,|\sin(n+1)\theta/2\,|] \qquad (n+1 \text{ even})$$

on $|\,z\,| \leq 1$ if q is within J. Thus (b) is established.

Now suppose $T(q) = b$ is a regular point of B. Let xby be an arc in B consisting wholly of regular points. Referring back to the proofs of (ii) and (iii) in § 2, we see that since $T^{-1}(xby)$ is a finite graph we can choose the simple closed curve X in those proofs so that

$$T^{-1}(xby)\cdot(R' + Z) = qx_1 + qx_2 + \cdots + qx_{2k}$$

where (1) $q = b'$, (2) the points $x_1, x_2, x_3, \cdots, x_{2k}$ are cyclically ordered on Z, (3) $qx_1, qx_2, \cdots$ are simple arcs intersecting in only q and such that qx_1 maps topologically onto the arc bu of xby, qx_2 onto bv, qx_3 onto bu, and so on to qx_{2k} which maps topologically onto bv, where $xby\cdot\bar{R}$ is the arc ubv, (4) the arc x_1x_2 of Z maps topologically onto one of the arcs, say urv, of X from u to v, x_2x_3 maps topologically onto the other arc, say usv, of Z, x_3x_4 maps onto urv, x_4x_5 onto usv and so on to $x_{2k}x_1$ which maps topologically onto usv. Thus by (5.1) it follows that the closed 2-cells $\bar{R}_1, \bar{R}_2, \cdots, \bar{R}_{2k}$ in $\bar{R}'$ bounded by the simple closed curves $qx_i + x_ix_{i+1} + qx_{i+1}$ $(0 \leq i \leq 2k,\ 2k+1 \equiv 1)$ map topologically and alternately onto the two 2-cells in $\bar{R}$ bounded by the closed curves $ubv + urv$ and $ubv + usv$. Hence, on $\bar{R}'$, T is topologically equivalent to the transformation $w = z^k$ on $|\,z\,| \leq 1$.

UNIVERSITY OF VIRGINIA.

ON THE DISTRIBUTION FUNCTIONS OF ALMOST PERIODIC FUNCTIONS.*

By PHILIP HARTMAN, E. R. VAN KAMPEN and AUREL WINTNER.

Introduction. While it is known[1] that every almost periodic function $z(t)$, $-\infty < t < +\infty$, has an asymptotic distribution function σ, very little is known about sufficient conditions which, when imposed on $z(t)$, insure a preassigned degree of smoothness for the function σ. Everything that is known about the subject indicates that this problem of smoothness depends essentially on at least two factors, namely

(i) the smoothness properties of the given function $z(t)$;

(ii) the arithmetical structure of the sequence of Fourier exponents of $z(t)$.

Actually, it is not (i) that is needed so much but rather

(i bis) the local smoothness properties of the function Z which is a continuous function of the position on a finite or infinite dimensional torus which is associated with $z(t)$ in the usual way.[2]

In fact, a condition of the type (i) is weaker than a condition of the type (i bis). For Z is constructed from the given function $z(t)$ by using the limiting process of the Kronecker-Weyl approximation theorem; and nothing seems to be known[3] about which, if any, of the smoothness properties of z (e. g., differentiability of a given degree or analyticity) are transplanted by this process into corresponding, or somewhat weaker, smoothness properties of the function Z on the torus. As far as the factor (ii) is concerned, it is known[4] that the situation for the smoothness of σ is most favorable in case that the Fourier exponents of $z(t)$ are linearly independent or such that $z(t)$ is of the form

$$z(t) = \Sigma\, z_\nu(\lambda_\nu t), \tag{1}$$

where the $z_\nu(t)$ are continuous functions of a fixed period T and the λ_ν are linearly independent. Correspondingly, it is to be expected that the chances

---

* Received January 31, 1938.

[1] Wintner [10], [11]; Haviland [3]. For a comprehensive account of a general theory, cf. Jessen and Wintner [6].

[2] Bohr [1].

[3] Cf. Wintner [13].

[4] Wintner [12]; Jessen and Wintner [6]; Kershner and Wintner [8]; van Kampen and Wintner [7].

for the smoothness of σ are least favorable[5] in case $z(t)$ is limit periodic, i. e., of the form

$$z(t)=\sum_{\nu=1}^{\infty} z_\nu(r_\nu t), \tag{2}$$

where the $z_\nu(t)\not\equiv z_0(0)$ are continuous functions of some common period and the r_ν are positive rational numbers such that $\liminf r_\nu = 0$.

The object of the present paper is to fill somewhat the gap between these two extreme cases. The method to be applied will be an extension of the one recently[6] applied in the most favorable case (1). From the point of view of the factor (ii), the smoothness problem of σ in case of an arbitrary trigonometric polynomial, i. e., of a finite number of Fourier exponents, is hardly different from the case of any almost periodic function whose Fourier exponents are generated by a sequence of linearly independent numbers (or more, generally, moduli). Correspondingly, when proving smoothness properties of σ, it will be a methodically unimportant simplification to assume that Z is a function on a finite dimensional torus.[7]

The results imply, in particular, that if $z(t)$ is a trigonometric polynomial, then its distribution function σ is absolutely continuous on the spectrum of σ. In particular, any non-constant trigonometric polynomial $x+iy=z(t)$ maps the t-axis on a connected set whose closure is either a finite set of analytic arcs or the closure of a two-dimensional open set which is bordered by a finite number of analytic arcs; and σ is necessarily absolutely continuous with a density which is always an analytic, though not necessarily regular analytic, function of the position on the spectrum, and which need not be one and the same analytic function on the entire spectrum. While this case of a trigonometric polynomial seems to be quite harmless, it is, as a fact, hardly easier than the general case to be considered.

As far as the factor (i bis) is concerned, the condition imposed on the function Z of the position on the finite dimensional torus is that either Z is regular analytic or that Z has continuous partial derivatives of order k, where $1\leqq k\leqq\infty$. In the second case, the results are similar to, although from the geometrical point of view more complicated than, those mentioned for the case of a trigonometric polynomial.

1. The open set *R*. Let $Z=Z(\theta_1,\cdots,\theta_n)$ be a continuous function of the position on the torus

---

[5] Cf. Bohr [2], where, however, the terms of the series (2) are not free of constant stretches.

[6] van Kampen and Wintner [7].

[7] As to a theory on an infinite dimensional torus, cf., Jessen [4], [5].

$$(3) \qquad \Theta: \quad 0 \leqq \theta_j < 1; \; j = 1, \cdots, n, \qquad (1 \leqq n < +\infty),$$

which is obtained from the n-dimensional Cartesian θ-space by reduction to modulus 1; and let T denote the corresponding mapping of Θ on the complex plane, so that

$$(4) \qquad T: \quad x + iy \equiv z = Z(\theta_1, \cdots, \theta_n).$$

By $T(\Lambda)$ will be denoted the image of a subset Λ of Θ under the continuous transformation T. While the continuous mapping T of Θ on $T(\Theta)$ is not, in general, topological, the set of all those points of Θ whose T-image is in a set F will be denoted by $T^{-1}(F)$, so that a subset Λ of Θ may be a proper subset of $T^{-1}(T(\Lambda))$. By meas Λ and $\mu(F)$ will be denoted the ordinary n-dimensional Lebesgue measure on Θ and the 2-dimensional Lebesgue measure in the (x, y)-plane, respectively.

It will be assumed that

(I) the function (4) is of class C_k, $k \geqq 1$, where C_k for $k \geqq 1$ is the class of functions for which all continuous partial derivatives of order k exist (and C_0 is the class of continuous functions).

(II) the mapping (4) has the property that meas $T^{-1}(P) = 0$ for every point P of the (x, y)-plane.

Condition (II) excludes the case that Z is constant on an open subset of Θ. Even if the function (4) has partial derivatives of arbitrarily high order and is nowhere constant, it is quite possible that condition (II) is not satisfied; for $T^{-1}(P)$ can be, for particular points P, a nowhere dense perfect set of positive measure.

Let $\lambda_1, \cdots, \lambda_n$ be n linearly independent real numbers, and $z(t)$ the almost periodic function

$$(5) \quad x(t) + iy(t) \equiv z(t) = Z(\lambda_1 t, \cdots, \lambda_n t), \quad (-\infty < t < +\infty).$$

Let E denote a Borel set in the (x, y)-plane, and $\sigma(E)$ the asymptotic distribution function of $z(t)$. It is clear from the Kronecker-Weyl approximation theorem that

$$(6) \qquad \sigma(E) = \text{meas}\, T^{-1}(E).$$

Obviously, condition (II) is equivalent to the assumption that $\sigma(E)$ has no point spectrum, i.e., that $\sigma(E) = 0$ whenever E consists of a single point. Since T is continuous, it is clear from (6) that the spectrum of $\sigma(E)$ is the set $T(\Theta)$.

Let Ω denote the set of those points of Θ at which the matrix

$$\begin{pmatrix} X_{\theta_1} & \cdots & X_{\theta_n} \\ Y_{\theta_1} & \cdots & Y_{\theta_n} \end{pmatrix}, \tag{7}$$

formed by the partial derivatives of the real and imaginary parts of (4), is not of rank 2; so that Ω consists of the zeros $(\theta_1, \cdots, \theta_n)$ of the Jacobian square sum

$$\Delta \equiv \Delta(\theta_1, \cdots, \theta_n) = \sum_{1 \leqq j < l \leqq n} \begin{vmatrix} X_{\theta_j} & X_{\theta_l} \\ Y_{\theta_j} & Y_{\theta_l} \end{vmatrix}^2. \tag{8}$$

Since the function (8) is continuous on Θ, the set Ω is closed. Let R denote the complement,

$$R = T(\Theta) - T(\Omega), \tag{9}$$

of the T-image of Ω with respect to the spectrum $T(\Theta)$, so that R is contained in, but is not necessarily identical with, $T(\Theta - \Omega)$.

It is easy to see that the set R, which may be empty, is always open. In fact, if P_0: (x_0, y_0) is a point of $T(\Theta - \Omega)$, then $P_0 = T(\Pi_0)$ for at least one point Π_0: $(\theta_1{}^0, \cdots, \theta_n{}^0)$ of the subset $\Theta - \Omega$ of Θ. Hence, the function (8) does not vanish at $(\theta_1{}^0, \cdots, \theta_n{}^0)$; in other words, at least one of the two-rowed Jacobian determinants occurring in (8), say the one for which $j = 1$, $l = 2$, does not vanish at $(\theta_1{}^0, \cdots, \theta_n{}^0)$. Thus, a small vicinity of the point $(\theta_1, \theta_2) = (\theta_1{}^0, \theta_2{}^0)$ of the (θ_1, θ_2)-subspace $\theta_3 = \theta_3{}^0, \cdots, \theta_n = \theta_n{}^0$ of Θ is mapped by T on a vicinity of P_0: (x_0, y_0) in a topological way, so that $T(\Theta - \Omega)$ is an open set. Obviously, $T(\Omega)$ is a closed set. Hence, R is an open set in the (x, y)-plane.

2. The distribution function on *R*. It will be shown that if R is not empty, the completely additive set function $\sigma(E)$ is absolutely continuous on R and has there a density of class C_{k-1}, i. e., there exists on R a (non-negative) function $\delta = \delta(x, y)$ of class C_{k-1} such that

$$\sigma(E) = \int_E \delta(x, y)\, dx dy \text{ for every } E \subset R. \tag{10}$$

First, if P_0: (x_0, y_0) is any point of R, every point of the closed set $T^{-1}(P_0)$ is contained in the open set $\Theta - \Omega$ on which the rank of the matrix (7) is 2. Hence, $T^{-1}(P_0)$ is not only a closed set, but it is also an $(n - 2)$-dimensional manifold, with the property that the common part of $T^{-1}(P_0)$ and of any sufficiently small sphere contained in Θ is a connected set. Since Θ is compact, it follows that $T^{-1}(P_0)$ consists of a finite number, say $m = m(P_0)$, of mutually disjoint, connected closed $(n - 2)$-dimensional manifolds

$N_1, \cdots, N_m$. Since P_0 is any point of the open set R, the compactness of Θ also implies the existence of a sufficiently small $\epsilon > 0$ such that the closure $\bar{C}$ of the circle C defined by

$$C: \qquad PP_0^2 = (x - x_0)^2 + (y - y_0)^2 < \epsilon^2$$

is contained in R and $m(P) = m(P_0)$ holds for every point $P: (x, y)$ of the closed set $\bar{C}$. The T^{-1}-image of C consists of m mutually disjoint, connected open sets on Θ. If $\Lambda_1, \cdots, \Lambda_m$ denote these open sets, then every Λ_j is a T^{-1}-image of C and every point of $T^{-1}(C)$ is contained in $\Lambda_1 + \cdots + \Lambda_m$. It will be supposed that the enumeration has been chosen in such a way that $\Lambda_q \supset N_q$, $q = 1, \cdots, m$.

It is clear from the definition of N_q that there exists in Θ a finite number of spheres, say p_q, such that these p_q spheres cover N_q and the common part $\Gamma_{qr}{}^{n-2}$ of N_q and the r-th sphere can be parametrized by functions of class C_k defined on an $(n-2)$-dimensional sphere.

Suppose now that $k \geqq 2$ in the assumption (I) of section 2 and that[8] the dimension number $n \geqq 3$. Then if $\Pi_0: (\theta_1{}^0, \cdots, \theta_n{}^0)$ is any point on N_q and if $M_0 = M(\theta_1{}^0, \cdots, \theta_n{}^0)$ denotes the (2-dimensional) plane normal to N_q at Π_0, then there exists an $\eta > 0$ (independent of Π_0) such that the η-vicinity of Π_0 on M_0 does not contain points of any other normal plane $M(\theta_1, \cdots, \theta_n)$ for any point $\Pi: (\theta_1, \cdots, \theta_n) \neq \Pi_0$ on any N_j, $j = 1, \cdots, m$. Furthermore, this vicinity of Π on M is the topological map of a neighborhood of $P_0: (x_0, y_0)$ under the transformation T; and this topological correspondence is given by functions of class C_k.

Let $\epsilon > 0$ be so small that the T^{-1}-image of the circle C, namely $\Lambda_1, \cdots, \Lambda_m$, is such that all points of the open set Λ_q are within a distance η of the set N_q. It is clear from the above considerations that Λ_q may be considered to be a "cylindrical tube" obtained by taking the image of C on all normal planes $M(\theta_1, \cdots, \theta_n)$ as the point $(\theta_1, \cdots, \theta_n)$ varies over N_q. Let $\Gamma_{qr}{}^n$ $(q = 1, \cdots, m;\ r = 1, \cdots, p_q)$ denote the n-dimensional set consisting of images of C on all normal planes $M(\theta_1, \cdots, \theta_n)$ as $(\theta_1, \cdots, \theta_n)$ varies over $\Gamma_{qr}{}^{n-2}$.

Now there exist functions of class C_k,

$$\theta_j = \vartheta_j{}^{qr}(x, y, \alpha_1, \cdots, \alpha_{n-2}), \tag{11}$$

$(j = 1, \cdots, n;\ q = 1, \cdots, m;\ r = 1, \cdots, p_q)$, defined on the product space of the $(n-2)$-dimensional sphere A: $\sum_{j=1}^{n-2} \alpha^2_j < 1$ and the set C on the (x, y)-plane, so that (11) is a parametrization of $\Gamma_{qr}{}^n$ for fixed q, r. In fact, for fixed q, r and fixed $x = x_0$, $y = y_0$, the functions (11) denote a parametrization of $\Gamma_{qr}{}^{n-2}$ (the same, of course, is true if (x_0, y_0) is replaced by an arbitrary point of C and $\Gamma_{qr}{}^{n-2}$ by the corresponding set of Θ); also, these functions (11) give, for fixed q, r and fixed $(\alpha_1, \cdots, \alpha_{n-2})$, a parametrization of the map of the circle C on the corresponding normal plane M. The Jacobian

---

[8] Cf. E. R. van Kampen and A. Wintner, [7], pp. 181-182 for a case analogous to $n = 2$.

$$J^{qr}(x, y, \alpha_1, \cdots, \alpha_{n-2}) = \frac{\partial(\vartheta_1{}^{qr}, \cdots, \vartheta_n{}^{qr})}{\partial(x, y, \alpha_1, \cdots, \alpha_{n-2})}$$

does not vanish on the product space of A and C.

It is clear from the above analysis of the mapping T^{-1} in a neighborhood of P_0, that (10) is satisfied for every open E (hence for every Borel set) in the ϵ-vicinity of P_0, if one puts

$$(12) \quad \delta(x, y) = \sum_{q=1}^{m} \sum_{r=1}^{p_q} \int \cdots \int |J^{qr}(x, y, \alpha_1, \cdots, \alpha_{n-2})| \, d\alpha_1 \cdots d\alpha_{n-2}$$

where the integral of $|J^{qr}|$ extends over that part of the sphere A which corresponds to points of $\Gamma_{qr}{}^n$ not contained in $\Gamma_{qj}{}^n$, $j = 1, \cdots, r-1$. This function (12) is of class C_{k-1}. Since P_0 is an arbitrary point of R, the proof is complete for $k \geqq 2$.

Obvious modifications of this proof assure the result for the case that the function (4) is of class C_1.

3. A criterion for the absolute continuity of $\sigma(E)$. If one does not introduce conditions in addition to those assumed so far, one cannot state that the asymptotic distribution function (6) of (5) is absolutely continuous, i. e., that there exists a (non-negative) measurable function $\delta(x, y)$ such that

$$(13) \qquad \sigma(E) = \int\int_E \delta(x, y)\, dxdy$$

holds not only for every Borel set E contained in the open (and possibly empty) set R but for every Borel set E contained in the spectrum of σ. Since $T(\Theta)$ is the spectrum, it is clear that the distribution function σ is absolutely continuous if

$$(14) \qquad \sigma(T(\Omega)) = 0.$$

It will be shown that if Ω is a zero set on Θ, i. e., if

$$(15) \qquad \text{meas}\, \Omega = 0,$$

then $\sigma(E)$ is an absolutely continuous distribution function and that, in addition, the measure $\mu(R)$ of the open subset (9) of the spectrum $T(\Theta)$ is identical with $\mu(T(\Theta))$. Since (10) was shown to hold for a continuous function $\delta(x,y)$ on R, it follows that if E is any closed rectangle in the (x,y)-plane, then the Lebesgue integral (13) is a proper or possibly improper Riemann integral according as E does not or does contain points of the boundary $T(\Omega)$ of the open set R.

First, it is clear from (6) that the statement (14) is equivalent to

$$\text{meas}\, T^{-1}(T(\Omega)) = 0, \tag{16}$$

where $T^{-1}(T(\Omega))$ is not to be confused with Ω. On the other hand, the assumption (15) implies, for every $\epsilon > 0$, the existence of an open set Γ_ϵ in Θ such that $\Omega \subset \Gamma_\epsilon$ and $\text{meas}\, \Gamma_\epsilon < \epsilon$. Hence, in order to prove (16), it is sufficient to show that the common part of $T^{-1}(T(\Omega))$ and $\Theta - \Gamma_\epsilon$ is a zero set on Θ. Since the set $\Theta - \Gamma_\epsilon$ is closed and is contained in the open set $\Theta - \Omega$ on which the function (8) does not vanish, it follows by arguments similar to those used in section 2 that it is sufficient to prove the relation $\mu(T(\Omega)) = 0$. But this relation is obvious from (8) and the definition of Ω.

This proves that (15) implies (14) and also that $\mu(R) = \mu(T(\Theta))$, the last relation being, in view of (9), equivalent to $\mu(T(\Omega)) = 0$.

It may be mentioned, that instead of Lebesgue measure, one could have used Jordan content, since the zero sets involved are closed sets.

4. The analytic case. It is seen from the proof of (10), (11) and (12) that if (4) is a regular analytic function of the n real variables $\theta_1, \cdots, \theta_n$ on Θ, then $\delta(x,y)$ is a regular analytic function of the position on the open set R of the real (x,y)-plane (it is understood that R need not be connected). Furthermore, it is clear from the definition of Ω that, in the analytic case, Ω consists of a finite number of manifolds each if which has a dimension number less than the dimension number n of Θ, unless the function (8) vanishes identically on Θ. It follows, therefore, from section 3 that in the analytic case either (8) vanishes identically on Θ or the distribution function $\sigma(E)$ is absolutely continuous with a density $\delta(x,y)$ which is a regular analytic function of the real variables (x,y) on G_i, where $G_1, G_2, \cdots$ is a sequence of mutually disjoint connected open sets in the real (x,y)-plane and $R = \Sigma G_i$. This implies, in particular, that if Z is regular analytic, $\Delta \not\equiv 0$, and $\lambda_1, \cdots, \lambda_n$ are linearly independent, then the closure of the set of values z attained by the almost periodic function (5) for $-\infty < t < +\infty$ is the closure of an open set in the z-plane, i. e., such as to have no one-dimensional parts. For this geometrical restriction on the spectrum of σ is a necessary condition for the absolute continuity of σ.

The case of a trigonometric polynomial $z(t)$, as described in the introduction, affords, among other things, the simplification that one can always find a finite number of linearly independent exponents $\lambda_1, \cdots, \lambda_n$ and a regular analytic function Z on Θ such that (5) is satisfied.

It may be mentioned that if it is only known that the nowhere constant function (4) has continuous partial derivatives of arbitrarily high order, it is quite possible that the distribution function is not absolutely continuous, although the spectrum is two-dimensional or even a Jordan region. This holds even in the particular case (1) of convolutions.[9]

5. The case $\Delta \equiv 0$. There remains to be considered the case of a regular analytic mapping (4) such that $\Delta \equiv 0$ on Θ (a case illustrated by a real-valued Z). There are two cases possible according as the matrix (7) is or is not of rank 0 at every point of Θ. In the first case, the function (4) is a constant on Θ. In the second case, the rank of (7) is 1 at all points of Θ which do not belong to a set Ω_0 consisting of a finite number of manifolds each of which has a dimension number less than the dimension number n of Θ. In what follows, only the latter case will be considered.

Since all the two-rowed Jacobians formed by the elements of the matrix (7) vanish at all points of Θ, while at least one of the $2n$ partial derivatives occurring in (7) does not vanish at every point of $\Theta - \Omega_0$, it is easy to show that the image of the torus Θ under the analytic mapping (4) is a one-dimensional connected analytic manifold of finite arc length, or, more precisely, the spectrum $T(\Theta)$ of $\sigma(E)$ consists of a sequence of analytic arcs which can have singularities and which have a total finite arc length. In order to see this, no use need to be made of known general theorems concerning analytic mappings of the compact set Θ (theorems which apply also to the case considered in the previous section).

Since the spectrum of σ is one-dimensional, hence a zero set in the (x, y)-plane, the set function $\sigma(E)$ cannot be absolutely continuous in the sense of (13). Correspondingly, it is plausible to replace the definition (13) of absolute continuity by the requirement that if s is the local length parameter on $T(\Theta)$, then

$$\sigma(S) = \int_S \delta(s)\, ds \tag{13 bis}$$

for every open arc S (or for every Borel set S) on $T(\Theta)$ and for a suitable

---

[9] In this connection, cf. Kershner [9].

(non-negative) function $\delta(s)$ of the position s on $T(\circledast)$. Then an obvious adaptation of the considerations of sections 2, 3 shows that the distribution function $\sigma(E)$ is an absolutely continuous on $T(\circledast)$ in this sense, and that $T(\circledast)$ contains a sequence of mutually disjoint open arcs $A_1, A_2, \cdots$ such that $\delta(s)$ is regular analytic on every A_i and $\tau(T(\circledast) - \Sigma A_i) = 0$, where $\tau(S)$ denotes the s-measure defined by

$$\tau(S) = \int_S ds.$$

Actually, there are only a finite number of A_i and, correspondingly, only a finite number of G_i in the previous section.

6. Dependence of the distribution function on the moduli. Assuming only that the function (4) is continuous on $\circledast$, denoting by λ the vector formed by the n real numbers $(\lambda_1, \cdots, \lambda_n)$, and by $\sigma_\lambda(E)$ the distribution function of the almost periodic function (5), the distribution function (6), which is independent of λ, is identical with $\sigma_\lambda(E)$ whenever the n components of λ are linearly independent. On the other hand, $\sigma_\lambda(E)$ depends on λ in a rather unstable way, if one does not require that the n components of λ be linearly independent. Nevertheless, the instability referred to does not appear at those points of the λ-space at which the n components of λ are linearly independent.

In order to formulate this statement in a precise manner, let L_ν, where $\nu = 0, 1, \cdots, n$, denote the set of those points of the n-dimensional vector space of λ at which there exist exactly ν dependencies between the n components of λ, these dependencies being homogeneous, linear, and having integral coefficients. Thus, L_n consists of the single point which represents the origin of the λ-space, while L_0 is a set of points λ for which $\sigma_\lambda(E)$ is given by (6). The complement, $L^* = L_1 + \cdots + L_n$, of L_0 is everywhere dense in the λ-space, but is contained in a dense sequence of hyperplanes. If the function (4) of n variables is not a continuous function of the position on a torus whose dimension number is less than n, it is easy to see that the distribution function $\sigma_\lambda(E)$, considered as a functional on the λ-space, $L_0 + L^*$, is discontinuous at every point λ of L^*.

Now, the statement is that the functional of λ is continuous at every point of L_0, although the set L^* of the discontinuity points is everywhere dense in the λ-space. In other words, $\sigma_\lambda \rightarrow \sigma_{\lambda_0}$ when λ tends on $L_0 + L^*$ to an arbitrary point λ_0 on L_0. Since $\sigma_\lambda(E)$ is independent of λ on L_0, it is sufficient to prove that $\sigma_\lambda \rightarrow \sigma_{\lambda_0}$ when λ tends on L^* to an arbitrary point λ_0 on L_0. But this follows by inspection of the existence proof (cf., Wintner [11]; Haviland [3]) of the asymptotic distribution function σ_λ of the almost periodic function (5).

THE JOHNS HOPKINS UNIVERSITY.

BIBLIOGRAPHY.

[1] H. Bohr, "Zur Theorie der fastperiodischen Funktionen, II," *Acta Mathematica*, vol. 46 (1925), pp. 101-214.

[2] H. Bohr, "Kleinere Beiträge zur Theorie der fastperiodischen Funktionen, II," *Danske Videnskabernes Selskab, Mathematisk-Fysiske Meddelelser*, vol. 10 (1930), no. 6, pp. 12-17.

[3] E. K. Haviland, "On statistical methods in the theory of almost periodic functions," *Proceedings of the National Academy of Sciences*, vol. 19 (1933), pp. 549-555.

[4] B. Jessen, *Bidrag til Integralteorien for Funktioner af uendelig mange Variable*, Copenhagen (1930).

[5] B. Jessen, "The theory of integration in a space of an infinite number of dimensions," *Acta Mathematica*, vol. 63 (1934), pp. 249-323.

[6] B. Jessen and A. Wintner, "Distribution functions and the Riemann zeta function," *Transactions of the American Mathematical Society*, vol. 38 (1935), pp. 48-88.

[7] E. R. van Kampen and A. Wintner, "Convolutions of distributions on convex curves and the Riemann zeta function," *American Journal of Mathematics*, vol. 59 (1937), pp. 175-214.

[8] R. Kershner and A. Wintner, "On the asymptotic distribution of almost periodic functions with linearly independent frequencies," *American Journal of Mathematics*, vol. 58 (1936), pp. 91-94.

[9] R. Kershner, "On the addition of convex curves, II," *American Journal of Mathematics*, vol. 59 (1937), pp. 423-426.

[10] A. Wintner, *Spektraltheorie der unendlichen Matrizen*, Leipzig (1929).

[11] A. Wintner, "Über die statistische Unabhängigkeit der asymptotischen Verteilungsfunktionen inkommensurabler Partialschwingungen," *Mathematische Zeitschrift*, vol. 36 (1933), pp. 618-629.

[12] A. Wintner, "Upon a statistical method in the theory of diophantine approximations," *American Journal of Mathematics*, vol. 55 (1933), pp. 309-331.

[13] A. Wintner, "Liouville systems and almost periodic functions," *American Journal of Mathematics*, vol. 60 (1938), pp. 463-472.

THE FOURIER COEFFICIENTS OF THE MODULAR INVARIANT $J(\tau)$.*

By Hans Rademacher.

1. Recently Dr. Zuckerman and I have developed general formulae for the Fourier coefficients of modular forms of positive dimensions.[1] We remarked at the end of our paper that the series obtained would be convergent also for forms of dimension zero, i. e. for *modular functions*, among which $J(\tau)$ can be regarded as fundamental. The question arises whether the formally constructed series for the coefficients of $J(\tau)$ actually represents them.

The solution of this problem requires a thorough revision of our method, since we had essentially made use of the positivity of the dimension of the modular forms. A method due to Kloosterman[2] and later extended by Estermann[3] gives the clue. Kloostermann's method consists of two devices: first it improves the estimate of certain sums $A_k(n)$ of roots of unity from the trivial one

$$|A_k(n)| \leqq k$$

to

$$|A_k(n)| \leqq Ck^{1-\beta+\epsilon} \cdot (k, n)^{\beta}, \tag{1.1}$$

where β, according to results of Salié[4] and Davenport[5] can be taken as $\beta = \frac{1}{3}$. Secondly it collects the Farey arcs $\xi_{h,k}$ belonging to the same k and treats the resulting sum as a whole instead of estimating the summands separately. Both of these expedients will be used in the following paper.

---

* Received February 20, 1938.

[1] "On the Fourier coefficients of certain modular forms of positive dimension," to be published in the *Annals of Mathematics*.

[2] H. D. Kloosterman, "Asymptotische Formeln für die Fourierkoeffizienten ganzer Modulformen," *Abhandlungen Hamburg. Math. Seminar*, vol. 5 (1927), pp. 337-352.

[3] T. Estermann, "Vereinfachter Beweis eines Satzes von Kloosterman," *ibid.*, vol. 7 (1929), pp. 82-98.

[4] H. Salié, "Zur Abschätzung der Fourierkoeffizienten ganzer Modulformen," *Mathematische Zeitschrift*, vol. 36 (1933), pp. 263-278.

[5] H. Davenport, "On certain exponential sums," *Journal f. d. reine u. angew. Mathematik*, vol. 169 (1933), pp. 158-176.

2. Our problem is to investigate the coefficients of the expansion

$$12^3 J(\tau) = e^{-2\pi i\tau} + \sum_{n=0}^{\infty} c_n e^{2\pi i n\tau} = f(e^{2\pi i\tau}), \tag{2.1}$$

where $J(\tau)$ is defined by means of the modular forms $g_2(\omega_1, \omega_2)$ and $g_3(\omega_1, \omega_2)$ as

$$J(\tau) = \frac{g_2{}^3(1,\tau)}{g_2{}^3(1,\tau) - 27 g_3{}^2(1,\tau)} = \frac{g_2{}^3(1,\tau)}{\Delta(1,\tau)}, \qquad \Im(\tau) > 0. \tag{2.2}$$

A consequence of (2.2) is, by the way, the formula[6]

$$12^3 J(\tau) = \frac{\left\{1 + 240 \sum\limits_{m=1}^{\infty} \dfrac{m^3 x^m}{1 - x^m}\right\}^3}{x \left\{\sum\limits_{\lambda=-\infty}^{+\infty} (-1)^\lambda x^{\frac{\lambda(3\lambda-1)}{2}}\right\}^{24}}, \qquad x = e^{2\pi i\tau}, \tag{2.3}$$

which shows that the coefficients c_n in (2.1) are integers. From (2.2) we infer the invariance of $J(\tau)$ with respect to the transformations of the full modular group:

$$J\left(\frac{a\tau + b}{c\tau + d}\right) = J(\tau). \tag{2.4}$$

We shall see that the equations (2.1) and (2.4) will completely suffice for the determination of the coefficients c_n, provided $n \geqq 1$. For

$$\tau = \frac{iz}{k} + \frac{h}{k}, \qquad \Re(z) > 0,$$

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} h' & -\dfrac{hh'+1}{k} \\ k & -h \end{pmatrix}$$

with

$$hh' \equiv -1 \pmod{k}, \tag{2.5}$$

equation (2.4) goes over into

$$J\left(\frac{iz}{k} + \frac{h}{k}\right) = J\left(\frac{i}{kz} + \frac{h'}{k}\right)$$

or, in the notation of (2.1),

$$f\left(e^{-\frac{2\pi z}{k} + \frac{2\pi i h}{k}}\right) = f\left(e^{-\frac{2\pi}{kz} + \frac{2\pi i h'}{k}}\right). \tag{2.6}$$

3. From (2.1) we obtain

$$c_n = \frac{1}{2\pi i} \int_C \frac{f(x)}{x^{n+1}}\, dx = \sum_{\substack{h,k \\ 0 \leqq h < k \leqq N}}' \frac{1}{2\pi i} \int_{\xi_{h,k}} \frac{f(x)}{x^{n+1}}\, dx,^7$$

---

[6] Klein–Fricke, *Vorlesungen über die Theorie der Modulfunktionen*, vol. I, p. 154.

[7] Σ' means here and subsequently that h runs over integers prime to k.

where the $\xi_{h,k}$ may be the Farey arcs of order N of the circle C

$$|x| = e^{-2\pi N^{-2}}.$$

If we introduce on $\xi_{h,k}$ the new variable ϕ through

$$x = \exp(-2\pi N^{-2} + \frac{2\pi ih}{k} + 2\pi i\phi)$$

we get

$$c_n = \sum_{\substack{h,k \\ 0 \leqq h < k \leqq N}}' e^{-\frac{2\pi inh}{k}} \int_{-\vartheta'_{h,k}}^{\vartheta''_{h,k}} f(e^{\frac{2\pi ih}{k} - 2\pi(N^{-2} - i\phi)}) e^{2\pi n(N^{-2} - i\phi)} d\phi. \tag{3.1}$$

For a later purpose we need here the determination of $\vartheta'_{h,k}$ and $\vartheta''_{h,k}$ in terms of h and k. In the Farey series of order N we consider the fraction h/k with its two neighbors:

$$\frac{h_1}{k_1} < \frac{h}{k} < \frac{h_2}{k_2}, \qquad k, k_1, k_2 \leqq N. \tag{3.2}$$

We have here

$$hk_1 - h_1k = 1, \qquad h_2k - hk_2 = 1,$$

and therefore

$$hk_1 \equiv 1 \pmod{k}, \qquad hk_2 \equiv -1 \pmod{k}$$

or, from (2.5),

$$k_1 \equiv -h', \qquad k_2 \equiv h' \pmod{k}. \tag{3.3}$$

The Farey segment around h/k is bounded by the mediants between the fractions (3.2)

$$\frac{h_1 + h}{k_1 + k}, \qquad \frac{h_2 + h}{k_2 + k}.$$

Since these mediants do not belong to the Farey series of order N we have

$$k_1 + k > N, \qquad k_2 + k > N,$$

which conditions, together with (3.2), enclose k_1 and k_2 in the intervals

$$N - k < k_1 \leqq N, \qquad N - k < k_2 \leqq N. \tag{3.4}$$

The formulae (3.3) and (3.4) determine k_1 and k_2 uniquely as functions of h and k. In particular we have

$$\vartheta'_{h,k} = \frac{1}{k(k_1 + k)}, \qquad \vartheta''_{h,k} = \frac{1}{k(k_2 + k)}. \tag{3.5}$$

4. In (3.1) we apply the transformation formula (2.6) and obtain

$$(4.1)\qquad c_n = \sum_{\substack{h,k\\ 0\leqq h<k\leqq N}}' e^{-\frac{2\pi i n h}{k}} \int_{-\vartheta'_{h,k}}^{\vartheta''_{h,k}} f\left(e^{\frac{2\pi i h'}{k}-\frac{2\pi}{k^2 w}}\right) e^{2\pi n w}\, d\phi$$

with the abbreviation

$$(4.2)\qquad w = N^{-2} - i\phi.$$

If we now write

$$(4.3)\qquad \begin{aligned} f(x) &= x^{-1} + D(x),\\ D(x) &= \sum_{m=0}^{\infty} c_m x^m \end{aligned}$$

we can accordingly split the expression (4.1) into two parts:

$$(4.4)\qquad c_n = Q(n) + R(n)$$

with

$$(4.41)\qquad Q(n) = \sum_{\substack{h,k\\ 0\leqq h<k\leqq N}}' e^{-\frac{2\pi i}{k}(nh+h')} \int_{-\vartheta'_{h,k}}^{\vartheta''_{h,k}} e^{\frac{2\pi}{k^2 w}+2\pi n w}\, d\phi,$$

$$(4.42)\qquad R(n) = \sum_{\substack{h,k\\ 0\leqq h<k\leqq N}}' e^{-\frac{2\pi i n h}{k}} \int_{-\vartheta'_{h,k}}^{\vartheta''_{h,k}} D\left(e^{\frac{2\pi i h'}{k}-\frac{2\pi}{k^2 w}}\right) e^{2\pi n w}\, d\phi.$$

In $Q(n)$, which we consider first, we divide the intervals of integration into three parts according to

$$-\vartheta'_{h,k} = -\frac{1}{k(k_1+k)} \leqq -\frac{1}{k(N+k)} < \frac{1}{k(N+k)} \leqq \frac{1}{k(k_2+k)} = \vartheta''_{h,k}$$

and get

$$\begin{aligned}(4.5)\qquad Q(n) &= \sum_{k=1}^{N} \sum_{h \bmod k}' e^{-\frac{2\pi i}{k}(nh+h')} \int_{-\frac{1}{k(N+k)}}^{\frac{1}{k(N+k)}}\\ &\quad + \sum_{k=1}^{N} \sum_{h \bmod k}' e^{-\frac{2\pi i}{k}(nh+h')} \int_{-\frac{1}{k(k_1+k)}}^{-\frac{1}{k(N+k)}}\\ &\quad + \sum_{k=1}^{N} \sum_{h \bmod k}' e^{-\frac{2\pi i}{k}(nh+h')} \int_{\frac{1}{k(N+k)}}^{\frac{1}{k(k_2+k)}}\\ &= Q_0(n) + Q_1(n) + Q_2(n),\end{aligned}$$

say. The integrand in all three integrals of (4.5) is the same as in (4.41).

5. In $Q_0(n)$ we can immediately perform the summation with respect to h since the integral is independent of h. Setting

$$A_k(n) = \sum_{h \bmod k}{}' e^{-\frac{2\pi i}{k}(nh+h')} \tag{5.1}$$

we get

$$Q_0(n) = \sum_{k=1}^{N} A_k(n) \int_{-\frac{1}{k(N+k)}}^{\frac{1}{k(N+k)}} e^{\frac{2\pi}{k^2 w}+2\pi n w}\, d\phi$$

or

$$Q_0(n) = \sum_{k=1}^{N} A_k(n) \frac{1}{i} \int_{N^{-2}-\frac{i}{k(N+k)}}^{N^{-2}+\frac{i}{k(N+k)}} e^{\frac{2\pi}{k^2 w}+2\pi n w}\, dw, \tag{5.2}$$

where we have introduced w from (4.2) as variable of integration. We remark further that $A_k(n)$ is a Kloosterman sum (cf. the references in § 1) and can therefore be estimated as

$$| A_k(n) | < C k^{2/3+\epsilon} \cdot (k, n)^{1/3}, \tag{5.3}$$

where (k, n) is the greatest common divisor of k and n. In the complex w-plane we consider now the closed rectangular path R with the four vertices

$$\pm N^{-2} \pm \frac{i}{k(N+k)}.$$

We take R as surrounding 0 in the positive sense. Then we have

$$\begin{aligned} Q_0(n) &= 2\pi \sum_{k=1}^{N} A_k(n) \frac{1}{2\pi i} \int_R e^{\frac{2\pi}{k^2 w}+2\pi n w}\, dw \\ &\quad - \frac{1}{i} \sum_{k=1}^{N} A_k(n) \left\{ \int_{N^{-2}+\frac{i}{k(N+k)}}^{-N^{-2}+\frac{i}{k(N+k)}} + \int_{-N^{-2}+\frac{i}{k(N+k)}}^{-N^{-2}-\frac{i}{k(N+k)}} + \int_{-N^{-2}-\frac{i}{k(N+k)}}^{N^{-2}-\frac{i}{k(N+k)}} \right\} \\ &= 2\pi \sum_{k=1}^{N} A_k(n) L_k(n) - \frac{1}{i} \sum_{k=1}^{N} A_k(n) \{J_1 + J_2 + J_3\}, \end{aligned} \tag{5.4}$$

say, where all four integrals have the same integrand.

For an estimation of J_1 and J_3 we observe that on their paths of integration we have

$$w = u \pm \frac{i}{k(N+k)},$$
$$-N^{-2} \leqq u \leqq N^{-2},$$
$$\Re(w) = u \leqq N^{-2},$$
$$\Re\left(\frac{1}{w}\right) = \frac{u}{u^2 + \dfrac{1}{k^2(N+k)^2}} < N^{-2}k^2(N+k)^2 \leqq 4k^2,$$

so that

$$\left| e^{\frac{2\pi}{k^2 w} + 2\pi n w} \right| \leqq e^{8\pi + 2\pi n N^{-2}}$$

and therefore

$$\left.\begin{matrix} |J_1| \\ |J_3| \end{matrix}\right\} \leqq 2N^{-2} e^{8\pi + 2\pi n N^{-2}}. \tag{5.5}$$

In J_2 we have

$$w = -N^{-2} + iv,$$
$$-\frac{1}{k(N+k)} \leqq v \leqq \frac{1}{k(N+k)},$$
$$\Re(w) = -N^{-2} < 0, \qquad \Re\left(\frac{1}{w}\right) = \frac{-N^{-2}}{N^{-4} + v^2} < 0,$$

hence

$$\left| e^{\frac{2\pi}{k^2 w} + 2\pi n w} \right| < 1$$

and therefore

$$|J_2| < \frac{2}{k(N+k)} < 2k^{-1}N^{-1}. \tag{5.6}$$

Combining (5.3), (5.5), and (5.6) we obtain

$$\sum_{k=1}^{N} A_k(n)\{J_1 + J_2 + J_3\} = O\left(e^{2\pi n N^{-2}} \sum_{k=1}^{N} k^{2/3+\epsilon}(n,k)^{1/3} k^{-1} N^{-1}\right)$$

and for $n \geqq 1$, which we assume from now on, we have $(n,k) \leqq n$ and hence

$$\sum_{k=1}^{N} A_k(n)\{J_1 + J_2 + J_3\} = O\left(e^{2\pi n N^{-2}} n^{1/3} N^{-1/3+\epsilon}\right). \tag{5.7}$$

Furthermore we have

$$L_k(n) = \frac{1}{2\pi i}\int_R e^{\frac{2\pi}{k^2w}+2\pi nw}\,dw$$

$$= \frac{1}{2\pi i}\int_R \sum_{\mu=0}^{\infty}\frac{\left(\frac{2\pi}{k^2w}\right)^{\mu}}{\mu!}\sum_{\nu=0}^{\infty}\frac{(2\pi nw)^{\nu}}{\nu!}\,dw$$

$$= \frac{1}{k\sqrt{n}}\sum_{\nu=0}^{\infty}\frac{\left(\frac{2\pi\sqrt{n}}{k}\right)^{2\nu+1}}{\nu!\,(\nu+1)!}$$

or

$$L_k(n) = \frac{1}{k\sqrt{n}}I_1\left(\frac{4\pi\sqrt{n}}{k}\right), \tag{5.8}$$

where $I_1(z)$ is the Bessel function of first order with purely imaginary argument. From (5.4), (5.7), (5.8) we deduce

$$Q_0(n) = \frac{2\pi}{\sqrt{n}}\sum_{k=1}^{N}\frac{A_k(n)}{k}\,I_1\left(\frac{4\pi\sqrt{n}}{k}\right) + O(e^{2\pi nN^{-2}}n^{1/3}N^{-1/3+\epsilon}). \tag{5.9}$$

6. We now turn our attention to $Q_1(n)$ and $Q_2(n)$ in (4.5), of which we discuss only $Q_2(n)$ in detail since $Q_1(n)$ can be treated in quite the same manner. We have from (4.5)

$$Q_2(n) = \sum_{k=1}^{N}\ \sideset{}{'}\sum_{h \bmod k} e^{-\frac{2\pi i}{k}(nh+h')}\int_{\frac{1}{k(N+k)}}^{\frac{1}{k(k_2+k)}} e^{\frac{2\pi}{k^2w}+2\pi nw}\,d\phi$$

$$= \sum_{k=1}^{N}\ \sideset{}{'}\sum_{h \bmod k} e^{-\frac{2\pi i}{k}(nh+h')}\sum_{l=k_2+k}^{N+k-1}\int_{\frac{1}{k(l+1)}}^{\frac{1}{kl}} e^{\frac{2\pi}{k^3w}+2\pi nw}\,d\phi,$$

$$Q_2(n) = \sum_{k=1}^{N}\sum_{l=N+1}^{N+k-1}\int_{\frac{1}{k(l+1)}}^{\frac{1}{kl}} e^{\frac{2\pi}{k^2w}+2\pi nw}\,d\phi \sideset{}{'}\sum_{\substack{h \bmod k\\ N<k_2+k\leqq l}} e^{-\frac{2\pi i}{k}(nh+h')}. \tag{6.1}$$

In the inner sum of the last expression the restriction imposed on k_2 means, in consequence of (3.3), a restriction of h' to an interval modulo k, which is equivalent to one interval or to two intervals in the range $0 \leqq h' < k$. Therefore the sum in question is an incomplete Kloosterman sum, for which we have the estimate [8]

---

[8] This estimate of the incomplete sum does not seem to appear explicitly in the

$$(6.2)\qquad \sum_{\substack{h \bmod k\\ N-k<k_2\leqq l-k}}{}' e^{-\frac{2\pi i}{k}(nh+h')} = O(k^{2/3+\epsilon}(n,k)^{1/3}) = O(k^{2/3+\epsilon}n^{1/3}).$$

In the integral in (6.1) we have

$$\begin{aligned}\Re\left(\frac{2\pi}{k^2w}+2\pi nw\right) &= \Re\left(\frac{2\pi}{k^2(N^{-2}-i\phi)}+2\pi n(N^{-2}-i\phi)\right)\\ &= 2\pi\left(\frac{N^{-2}}{k^2(N^{-4}+\phi^2)}+nN^{-2}\right)\leqq 2\pi\left(\frac{N^{-2}}{k^2N^{-4}+\dfrac{1}{(k+N)^2}}+nN^{-2}\right)\\ &< 2\pi\left(\left(\frac{k+N}{N}\right)^2+nN^{-2}\right)\leqq 8\pi+2\pi nN^{-2},\end{aligned}$$

and from this and (6.2) we get

$$\begin{aligned}Q_2(n) &= O\left(e^{2\pi nN^{-2}}n^{1/3}\sum_{k=1}^{N}\sum_{l=N+1}^{N+k-1}\left(\frac{1}{kl}-\frac{1}{k(l+1)}\right)k^{2/3+\epsilon}\right)\\ &= O\left(e^{2\pi nN^{-2}}n^{1/3}\sum_{k=1}^{N}\frac{1}{k^{1/3-\epsilon}N}\right),\end{aligned}$$

$$(6.3)\qquad Q_2(n) = O(e^{2\pi nN^{-2}}n^{1/3}N^{-1/3+\epsilon}).$$

Since a similar result is valid for $Q_1(n)$ we derive from (4.5), (5.9), and (6.3):

$$(6.4)\qquad Q(n) = \frac{2\pi}{\sqrt{n}}\sum_{k=1}^{N}\frac{A_k(n)}{k}I_1\left(\frac{4\pi\sqrt{n}}{k}\right)+O(e^{2\pi nN^{-2}}n^{1/3}N^{-1/3+\epsilon}).$$

7. From (4.42) and (4.3) we obtain

$$R(n) = \sum_{k=1}^{N}\sum_{h \bmod k}{}' e^{-\frac{2\pi inh}{k}}\int_{-\vartheta'_{h,k}}^{\vartheta''_{h,k}}\sum_{m=0}^{\infty}c_m e^{\frac{2\pi ih'm}{k}-\frac{2\pi m}{k^2w}}e^{2\pi nw}\,d\phi.$$

We decompose again the Farey segment $-\vartheta'_{h,k}\leqq\phi\leqq\vartheta''_{h,k}$ in the Kloosterman manner and have, after an interchange of the summations with respect to h and m,

---

literature. Incomplete sums for general k are given in Kloosterman's paper with a less precise estimate, and Davenport treats only the case k equal to a prime number, but with the precision (6.2). By the device, however, which Estermann uses *loc. cit.*, p. 94, or by the other one, which Davenport applies *loc. cit.*, pp. 173, 174, we can reduce the estimation of the incomplete sum to that of the complete sum. Thus we obtain (6.2) from Salié's estimate, *loc. cit.*, p. 264. We are for our purpose not particularly interested in the lowest possible value of the exponent of k in (6.2), as long as it is a constant less than 1.

$$
\begin{aligned}
(7.1)\quad R(n) &= \sum_{k=1}^{N}\sum_{m=0}^{\infty} c_m \int_{-\frac{1}{k(N+k)}}^{\frac{1}{k(N+k)}} e^{-\frac{2\pi m}{k^2 w}+2\pi n w}\,d\phi \sum_{h \bmod k}{}' e^{-\frac{2\pi i}{k}(nh-mh')} \\
&+ \sum_{k=1}^{N}\sum_{m=0}^{\infty} c_m \sum_{h \bmod k}{}' e^{-\frac{2\pi i}{k}(nh-mh')} \sum_{l=k_1+k}^{N+k-1} \int_{-\frac{1}{kl}}^{-\frac{1}{k(l+1)}} e^{-\frac{2\pi m}{k^2 w}+2\pi n w}\,d\phi \\
&+ \sum_{k=1}^{N}\sum_{m=0}^{\infty} c_m \sum_{h \bmod k}{}' e^{-\frac{2\pi i}{k}(nh-mh')} \sum_{l=k_2+k}^{N+k-1} \int_{\frac{1}{k(l+1)}}^{\frac{1}{kl}} e^{-\frac{2\pi m}{k^2 w}+2\pi n w}\,d\phi \\
&= S_1 + S_2 + S_3.
\end{aligned}
$$

In all integrals of (7.1) we have

$$
(7.2)\quad \Re\left(\frac{2\pi m}{k^2 w}\right) = \frac{2\pi m N^{-2}}{k^2(N^{-4}+\phi^2)} \geqq \frac{2\pi m}{k^2 N^{-2} + N^2 k^2 \vartheta^2_{h,k}} \geqq \frac{2\pi m}{1+1} = \pi m.
$$

The complete Kloosterman sum in S_1 admits of an estimate

$$
\sum_{h \bmod k}{}' e^{-\frac{2\pi i}{k}(nh-mh')} = O\left(k^{2/3+\epsilon}(n,k)^{1/3}\right) = O\left(k^{2/3+\epsilon} n^{1/3}\right)
$$

which holds uniformly in m. We obtain therefore

$$
\begin{aligned}
S_1 &= O\left(\sum_{k=1}^{N}\sum_{m=0}^{\infty} |\, c_m \,| \frac{2}{kN} e^{-\pi m + 2\pi n N^{-2}} k^{2/3+\epsilon} n^{1/3}\right) \\
&= O\left(e^{2\pi n N^{-2}} n^{1/3} N^{-1} \sum_{m=0}^{\infty} |\, c_m \,|\, e^{-\pi m} \sum_{k=1}^{N} k^{-1/3+\epsilon}\right),
\end{aligned}
$$

$$
(7.3)\quad S_1 = O\left(e^{2\pi n N^{-2}} n^{1/3} N^{-1/3+\epsilon}\right).
$$

The sums S_2 and S_3 are both of the same structure so that we need to treat only one of them. By interchanging the summations with respect to h and l we get

$$
S_2 = \sum_{k=1}^{N}\sum_{m=0}^{\infty} c_m \sum_{l=N+1}^{N+k-1} \int_{-\frac{1}{kl}}^{-\frac{1}{k(l+1)}} e^{-\frac{2\pi m}{k^2 w}+2\pi n w}\,d\phi \sum_{\substack{h \bmod k \\ N < k+k_1 \leqq l}}{}' e^{-\frac{2\pi i}{k}(nh-mh')}.
$$

The inner sum is an incomplete Kloosterman sum, for which we have, uniformly in m, the estimate

$$\sum_{\substack{h \bmod k \\ N-k<k_1\leqq l-k}}' e^{-\frac{2\pi i}{k}(nh-mh')} = O(k^{2/3+\epsilon}(n,k)^{1/3}) = O(k^{2/3+\epsilon}n^{1/3})$$

Therefore, taking note of (7.2), we get

$$S_2 = O\left(\sum_{k=1}^{N}\sum_{m=0}^{\infty} |\, c_m \,| \frac{1}{kN}\, e^{-\pi m+2\pi nN^{-2}} k^{2/3+\epsilon} n^{1/3}\right),$$

$$S_2 = O(e^{2\pi nN^{-2}} n^{1/3} N^{-1/3+\epsilon}). \tag{7.4}$$

From (7.1), (7.3), and (7.4) we infer

$$R(n) = O(e^{2\pi nN^{-2}} n^{1/3} N^{-1/3+\epsilon}) \tag{7.5}$$

and then from (4.4), (6.4), (7.5)

$$c_n = \frac{2\pi}{\sqrt{n}} \sum_{k=1}^{N} \frac{A_k(n)}{k} I_1\left(\frac{4\pi\sqrt{n}}{k}\right) + O(e^{2\pi nN^{-2}} n^{1/3} N^{-1/3+\epsilon}). \tag{7.6}$$

Now we keep here $n > 0$ fixed and let N tend to infinity. The error term then tends to zero. Thus we obtain our main result, which we state in the following

THEOREM. *In the Fourier expansion for the modular function* $J(\tau)$

$$12^3 J(\tau) = e^{-2\pi i t} + c_0 + \sum_{n=1}^{\infty} c_n e^{2\pi i n\tau} \tag{7.71}$$

the coefficients c_n, $n \geqq 1$, *are determined by the convergent series*

$$c_n = \frac{2\pi}{\sqrt{n}} \sum_{k=1}^{\infty} \frac{A_k(n)}{k} I_1\left(\frac{4\pi\sqrt{n}}{k}\right) \tag{7.72}$$

with

$$A_k(n) = \sum_{h \bmod k}' e^{-\frac{2\pi i}{k}(nh+h')}, \qquad hh' \equiv -1 \pmod{k}. \tag{7.73}$$

8. We had to exclude $n = 0$ in our discussion. This peculiarity is not caused by $\sqrt{n}$ appearing in the denominator in (7.6), as the computation of $L_k(n)$ in the lines preceding (5.8) shows that $n = 0$ is not exceptional in this respect:

$$\left[\frac{1}{k\sqrt{n}} I_1\left(\frac{4\pi\sqrt{n}}{k}\right)\right]_{n=0} = L_k(0) = \frac{2\pi}{k^2}. \tag{8.1}$$

The estimates of the incomplete Kloosterman sums, however, would break down for $n = 0$. By suitable examples it is easy to see that incomplete Kloosterman sums with $n = 0$ do not admit of a better general estimate than $O(k)$, which would, of course, not suffice in our reasonings. The series (7.72), on the other hand, will remain convergent for $n = 0$, but does not even accidentally represent the coefficient c_0, which can directly be obtained from (2.3) as

$$c_0 = 744. \tag{8.2}$$

Indeed, we get

$$A_k(0) = \sum_{h \bmod k}' e^{-\frac{2\pi i}{k}h'} = \mu(k)$$

with the Möbius symbol $\mu(k)$, and hence from (8.1) and (7.72)

$$2\pi \sum_{k=1}^{\infty} A_k(0) \frac{2\pi}{k^2} = 4\pi^2 \sum_{k=1}^{\infty} \frac{\mu(k)}{k^2} = 4\pi^2 \frac{1}{\zeta(2)} = 4\pi^2 \cdot \frac{6}{\pi^2} = 24,$$

different from (8.2). There is of course no reason to expect that c_0 might be found by our method, which makes use only of the behavior of $J(\tau)$ at $\tau = i\infty$, expressed by (2.1), and of the invariance stated in (2.4). Both properties remain obviously unchanged for any $J(\tau) + C$ instead of $J(\tau)$.

9. The coefficients c_n, which can be found from (2.3) by troublesome computations, which for higher n are practically inexecutable, do not seem to have attracted much attention before. All I could discover in the literature were, besides c_0, the two coefficients

$$c_1 = 196\,884, \qquad\qquad c_2 = 21\,493\,760.$$

Our convergent series (7.72) gives another approach to the actual computation of the c_n, which, as we know, have to be integers. Unfortunately, the convergence of (7.72) is rather slow, so that we should need quite a number of terms in order to get an error which is safely less than 1/2. Nevertheless, it is interesting to see that the first few terms of the series already furnish the bulk of the considerable amounts of those coefficients.

We get, for $n = 1$:

$$\begin{aligned}
2\pi \frac{A_1(1)}{1} I_1(4\pi) &= 196\,550.665 \\
2\pi \frac{A_2(1)}{2} I_1\left(\frac{4\pi}{2}\right) &= 250.822 \\
2\pi \frac{A_3(1)}{3} I_1\left(\frac{4\pi}{3}\right) &= 48.535 \\
2\pi \frac{A_4(1)}{4} I_1\left(\frac{4\pi}{4}\right) &= 14.110 \\
2\pi \frac{A_5(1)}{5} I_1\left(\frac{4\pi}{5}\right) &= 8.380 \\
\hline
\sum &= 196\,872.512,
\end{aligned}$$

and for $n = 2$:

$$\begin{aligned}
\frac{2\pi}{\sqrt{2}}\,\frac{A_1(2)}{1}\,I_1(4\pi\sqrt{2}) &= 21\,495\,869.279\\
\frac{2\pi}{\sqrt{2}}\,\frac{A_2(2)}{2}\,I_1\left(\frac{4\pi\sqrt{2}}{2}\right) &= -2\,054.739\\
\frac{2\pi}{\sqrt{2}}\,\frac{A_3(2)}{3}\,I_1\left(\frac{4\pi\sqrt{2}}{3}\right) &= -\quad 84.640\\
\frac{2\pi}{\sqrt{2}}\,\frac{A_4(2)}{4}\,I_1\left(\frac{4\pi\sqrt{2}}{4}\right) &= \quad 0.000\\
\frac{2\pi}{\sqrt{2}}\,\frac{A_5(2)}{5}\,I_1\left(\frac{4\pi\sqrt{2}}{5}\right) &= \quad 7.012\\
\sum &= \overline{21\,493\,736.912.}
\end{aligned}$$

These values are in error only by the comparatively small amounts of -11.488 and -23.088 respectively.

UNIVERSITY OF PENNSYLVANIA,
PHILADELPHIA, PA.

CAL CENT LIBRAR UNIVERSIT

FOURIER-STIELTJES TRANSFORMS AND SINGULAR INFINITE CONVOLUTIONS.*

By Norbert Wiener and Aurel Wintner.

1. Let $F(x)$, $-\pi \leqq x \leqq \pi$, be a continuous monotone function and

$$(1) \qquad c_n = c_n(F) = \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{-inx} dF(x) = \bar{c}_{-n} \qquad (n = 0, 1, 2, \cdots)$$

the sequence of its trigonometrical moments or Fourier-Stieltjes coefficients. If $F(x)$ is absolutely continuous, i. e., if

$$(2) \qquad F(x) = \int^{x} f(y) dy \qquad (-\pi \leqq x \leqq \pi)$$

holds for an L-integrable function f, then

$$(3) \qquad c_n = o(1)$$

by the Riemann-Lebesgue lemma. The converse is not true, since an example of Menchoff [8] shows that (3) may hold for a monotone function $F(x)$ which is singular, i. e., continuous and such that its derivative vanishes almost everywhere but $F(x) \neq \text{const.}$

Let an absolute constant κ^* be defined by means of a Dedekind cut as follows: For a given number κ, there do or do not exist singular monotone functions $F(x)$ which satisfy

$$(4) \qquad c_n = o(n^{-\kappa})$$

according as $\kappa < \kappa^*$ or $\kappa > \kappa^*$. Menchoff's result, when expressed in terms of κ^*, does not state more than that κ^* cannot be negative, a fact which is implied by $c_n = O(1)$ and is, therefore, clear from (1). Also a result of Kershner [6], which has improved on Menchoff's example (3), does not give more than $\kappa^* \geqq 0$. However, Littlewood [7] has recently proved by an ingenious, though rather involved, construction that $\kappa^* \neq 0$. Littlewood's proof of $\kappa^* > 0$ is a pure existence proof, since it seems to be hardly possible to prove by his method that $\kappa^* > 10^{-10}$, say. At any rate, Littlewood's result $\kappa^* > 0$ implies the existence of a critical κ-range within which (4) is incapable

---

* Received May 9, 1938.

of deciding whether or not $F(x)$ is absolutely continuous. In fact, a well-known remark of Lebesgue shows that the Riemann-Lebesgue lemma, i. e. (3), cannot be improved for the class of all absolutely continuous F (not even if f in (2) is restricted to be continuous and periodic).

On the other hand, the critical κ-range is bounded, i. e., $\kappa^* < +\infty$. Actually, $\kappa^* \leqq \frac{1}{2}$. In fact, if F is such that (4) holds for a $\kappa > \frac{1}{2}$, then $\Sigma \mid c_n \mid^2 < +\infty$, and so the Fischer-Riesz theorem assures that (2) holds for a function f which is of class L, since it is of class L^2.

2. Accordingly, $0 < \kappa^* \leqq \frac{1}{2}$. No better result (e. g., $10^{-10} \leqq \kappa^* \leqq \frac{1}{2} - 10^{-10}$) is known to us from the existing literature. But we shall prove that $\kappa^* = \frac{1}{2}$. In other words, we shall show that there exists for every $\epsilon > 0$ a singular monotone function $F = F_\epsilon$ such that

$$c_n = O(n^{-\frac{1}{2}+\epsilon}). \tag{5}$$

It seems to be a reasonable guess that if $\gamma_1, \gamma_2, \cdots$ are given positive numbers for which $\Sigma \gamma_n{}^2 = +\infty$, then there exists a singular monotone function F such that $c_n = O(\gamma_n)$ or $c_n = o(\gamma_n)$.

It will remain undecided whether or not the width, $\kappa^* = \frac{1}{2}$, of the critical κ-range is directly connected with the critical widths $\frac{1}{2}$ in the theory of ordinary Dirichlet series (although the method will be such as to prove (5) also when n in (1) is allowed to vary continuously). Actually, it will be seen that the absolute constant κ^* is the "same" $\frac{1}{2}$ which occurs in the standard asymptotic formulae of the type

$$J_0(u) = O(\mid u \mid^{-\frac{1}{2}}), \qquad J_0(u) = \Omega(\mid u \mid^{-\frac{1}{2}}).$$

In fact, the exponent $-\frac{1}{2}$ of (5) will be introduced by using van der Corput's lemma,

$$\mid \int_a^b \cos g(x)\,dx \mid \leqq \text{Const.} \min_{a \leqq x \leqq b} \mid g''(x) \mid^{-\frac{1}{2}}, \tag{6}$$

where $g(x)$, $a \leqq x \leqq b$, is any real function with a non-vanishing continuous second derivative $g''(x)$, and Const. is independent of a, b and g (cf. Kershner [5]).

The proof of $\kappa^* = \frac{1}{2}$ will consist of four steps:

(i) It turns out (§ 5) that, instead of proving the existence of a singular monotone $F(x) = F_\epsilon(x)$ for which the $c_n = c_n(F)$ are small in the sense of (5), it is sufficient to prove the existence of a singular monotone $G(x) = G_\epsilon(x)$ for which the $c_n = c_n(G)$ are small on the *average*, in the sense that

$$(7)\qquad \sum_{n=-N}^{N} |c_n| = O(N^{\epsilon}) \qquad (N \to +\infty).$$

(ii) It will be shown (§ **5**) that an $F(x)$, $-\pi \leqq x \leqq \pi$, which satisfies (5) can always be derived from a $G(x)$, $-\pi \leqq x \leqq \pi$, which satisfies (7), the passage from G to F consisting of an elementary one-to-one continuous transformation of the interval $-\pi \leqq x \leqq \pi$ into itself. It is only through this transformation that (5) becomes satisfied, since before this passage from G to F not even (3) is satisfied (cf. (29); § **7**). The effect of this transformation on the Fourier-Stieltjes coefficients may be estimated by means of the second mean value theorem, which introduces into (5) the exponent $-\frac{1}{2}$ of (6).

(iii) The theory of symmetric Bernoulli convolutions (cf. Jessen and Wintner [3], Kershner and Wintner [4], and Wintner [10], where further references are given) implies that the construction of a G which satisfies (7) can be based (§ **7**) on suitable estimations of the function

$$(8)\qquad L(u) = \prod_{k=1}^{\infty} \cos(3^{-k}u) \qquad (-\infty < u < +\infty)$$

for large $|u|$. Since $\cos v = 1 + O(v^2)$ as $v \to 0$, the product (8) defines an entire function of the exponential type. Actually, (8) is the Fourier-Stieltjes transform of the symmetric Bernoulli convolution which defines the classical Cantor function (cf. Carleman [1], pp. 223-226). Thus, while singular Bernoulli convolutions will occur, they will be used for quite another purpose than in the examples of Menchoff [8] and Kershner [6], examples which are in themselves singular Bernoulli convolutions and satisfy (3).

(iv) The remaining step, namely that estimate of (8) which was mentioned under (iii), will be made (§ **6**) by an adaptation of an enumeration process which is familiar in a similar, though easier, instance; cf. Wiener [9].

All of this will apply also for Fourier-Stieltjes transforms (§ **8**), instead of trigonometrical moments; so that in (1) the index n can be replaced by a continuous variable u.

3. Let $\xi = \xi(x)$, $-\pi \leqq x \leqq \pi$, be a strictly increasing continuous function such that $\xi(-\pi) = -\pi$, $\xi(\pi) = \pi$. Then, if $G(x)$, $-\pi \leqq x \leqq \pi$, is any continuous monotone function, so is the function $F(x)$, $-\pi \leqq x \leqq \pi$, which is defined by $G(x) = F(\xi(x))$, i. e., by $F(x) = G(\xi^{-1}(x))$. Clearly,

$$\int_{-\pi}^{\pi} e^{-imx} dF(x) = \int_{-\pi}^{\pi} e^{-imx} dG(\xi^{-1}(x)) = \int_{-\pi}^{\pi} e^{-im\xi(x)} dG(x).$$

Hence, on placing

$$(9) \qquad \lambda_{nm} = \frac{1}{2\pi}\int_{-\pi}^{\pi} e^{-im\xi(x)}e^{inx}dx, \qquad (m, n = 0, \pm 1, \cdots),$$

and developing the function $e^{-im\xi(x)}$, which is of bounded variation for every fixed m, into its Fourier series

$$e^{-im\xi(x)} = \sum_{n=-\infty}^{+\infty} \lambda_{nm}e^{-inx}, \qquad (m = 0, \pm 1, \cdots),$$

finally applying, for instance, a standard theorem of W. H. Young, one clearly obtains

$$(10) \qquad c_m(F) = \sum_{n=-\infty}^{+\infty} \lambda_{nm}c_n(G), \qquad (m = 0, \pm 1, \cdots),$$

if use is made of the notation (1).

Suppose, in particular, that the topological transformation $\xi = \xi(x)$ of the interval $[-\pi, \pi]$ into itself is such that the derivative of the function $\xi(x)$, or the distortion of the mapping $\xi = \xi(x)$, lies between two positive bounds. In this case it is clear that a monotone continuous function $G(x)$ is singular if and only if the same holds for the function $F(x) = G(\xi^{-1}(x))$.

4. Let

$$(11) \qquad \xi = \xi(x) = \tfrac{1}{4}(3x + \tfrac{1}{\pi}x^2 \operatorname{sgn} x), \quad \text{where} \quad -\pi \leqq x \leqq \pi.$$

This $\xi(x)$ is continuous, strictly monotone and such that $\xi(-\pi) = -\pi$, $\xi(\pi) = \pi$. The coefficients (9) belonging to the mapping (11) are

$$(12) \qquad \lambda_{nm} = \frac{1}{\pi}\int_{0}^{\pi} \cos\{\tfrac{1}{4}m(3x + \tfrac{1}{\pi}x^2) - nx\}dx.$$

Hence, from (6),

$$(13) \qquad |\lambda_{nm}| < \text{Const.}\, |m|^{-\frac{1}{2}} \text{ for all } n, m.$$

It will be shown that

$$(14) \qquad |\lambda_{nm}| < \text{Const.}\, n^{-2} \text{ whenever } |n| > 2|m|.$$

First, it is easily verified by differentiation with respect to x that if $|n| > 2|m|$, the function $\{\ \}$ whose cosine is the integrand of (12) is strictly monotone for $0 \leqq x \leqq \pi$. Hence, on placing $\{\ \} = t$ and

$$(15) \qquad f_{nm}(t) = |(\tfrac{3}{4}m - n)^2 + \tfrac{1}{\pi}mt|^{-\frac{1}{2}},$$

the integral (12) can be written as

$$\lambda_{nm} = \pm\tfrac{1}{\pi}\int_0^{(m-n)\pi} f_{nm}(t)\cos t\,dt.$$

Consequently, on applying partial integration and noting that $\frac{1}{\pi} < 1$,

$$|\lambda_{nm}| \leqq \Big|\int_0^{(m-n)\pi} f'_{nm}(t)\sin t\,dt\Big|, \text{ where } {}' = d/dt.$$

Now, it is easily verified from the assumption $|n| > 2|m|$ that the derivative, $f'_{nm}(t)$, of the function (15) is monotone between $t = 0$ and $t = (m - n)\pi$. Consequently, application of the second mean value theorem to the last integral gives

$$|\lambda_{nm}| \leqq 2\,|f'_{nm}(0)| + 2\,|f'_{nm}((m-n)\pi)|.$$

Hence, on calculating $f'_{nm}(0)$ and $f'_{nm}((m-n)\pi)$ from (15) and then using the assumption $|n| > 2|m|$, the appraisal (14) follows.

Incidentally, (13) also is a consequence of the second mean value theorem, since so is (6).

5. Writing (10) in the form

$$c_m(F) = \Big\{\sum_{n=-\infty}^{-2|m|-1} + \sum_{n=-2|m|}^{2|m|} + \sum_{n=2|m|+1}^{+\infty}\Big\}\lambda_{nm}c_n(G),$$

(13) and (14) imply that, as $m \rightarrow \pm\infty$,

$$c_m(F) = O\Big(\sum_{n=-\infty}^{-2|m|} n^{-2}\Big) + O\Big(\sum_{n=-2|m|}^{2|m|} |c_n(G)|\, m^{-\frac{1}{2}}\Big) + O\Big(\sum_{n=2|m|}^{+\infty} n^{-2}\Big),$$

since $|c_n(G)| \leqq c_0(G) = \text{const.}$, by (1). Accordingly,

$$c_m(F) = O(|m|^{-1}) + |m|^{-\frac{1}{2}}\,O\Big(\sum_{n=-2|m|}^{2|m|} |c_n(G)|\Big) \text{ as } m \rightarrow \pm\infty.$$

Hence, if $G(x)$ satisfies (7) for a fixed ϵ, then $F(x)$ satisfies (5) for the same ϵ. Furthermore, application of the last remark of § **3** to the mapping (11) shows that both monotone functions $F(x)$, $G(x)$ are singular if either of them is. This proves (i)-(ii), § **2**.

6. Next, (iv), § **2** will be established, by showing that there exists for every $\epsilon > 0$ a sufficiently large positive integer $p = p_\epsilon$ such that for the function (8) one has

$$(16)\qquad \sum_{n=-N}^{N} |L(2\pi n)|^p = O(N^\epsilon) \text{ as } N \to +\infty,$$

and even

$$(17)\qquad \sum_{n=-N}^{N} |L(\alpha n)|^p = O(N^\epsilon) \text{ uniformly for } \pi \leqq \alpha \leqq 2\pi, \text{ as } N \to +\infty.$$

First, if n is a positive integer, let $\nu = \nu_n$ denote the number of digits in the triadic representation of n; so that

$$(18)\qquad n = \delta_{\nu+1}\delta_\nu \cdots \delta_k \cdots \delta_2\delta_1, \qquad (\nu = \nu_n),$$

where every δ_k is either 0 or 1 or 2, while $\delta_\nu \neq 0$ and $\delta_{\nu+1} = 0$. Clearly, the fractional part, $\beta - [\beta]$, of $\beta = 3^{-k}n$ is

$$0.\, \delta_k\delta_{k-1} \cdots \delta_2\delta_1 \qquad (k = 1, \cdots, \nu_n).$$

Hence, on considering the six possible pairs

$$\{\delta_k; \delta_{k-1}\} = \{0;1\},\ \{1;0\},\ \{1;2\},\ \{2;1\},\ \{2;0\},\ \{0;2\},$$

it is easily verified that the fractional part of $3^{-k}n$ either lies between $1/9$ and $4/9$ or between $5/9$ and $8/9$. Consequently, in both cases,

$$|\cos(3^{-k}2\pi n)| \leqq \cos(\pi/9) < \exp(-1/100).$$

Since this holds for $k = 1, \cdots, \nu_n$, and since every factor of the product (8) lies between -1 and 1, it follows that

$$|L(2\pi n)| < \exp(-\mu_n/100),$$

where μ_n denotes the number of changes of the digits δ in the finite sequence (18), i. e., the number of those k for which $\delta_{k+1} \neq \delta_k$. Thus, if $p > 0$,

$$(19)\qquad |L(2\pi n)|^p < \exp(-p\mu_n/100).$$

Now, if m is a positive integer, the number of those among the first N positive integer n which have exactly m changes of digits in their triadic representation (18), i. e., the number of those n for which $1 \leqq n \leqq N$ and $\mu_n = m$, cannot exceed

$$(6\nu_N)^m/m!,$$

six being the number of distinct pairs of digits 0, 1, 2. It follows, therefore, from (19) that if the positive integer p is fixed and $N \to +\infty$, then

$$\sum_{n=1}^{N} |L(2\pi n)|^p = O\{\sum_{m=0}^{\infty} \exp(-pm/100) \cdot (6\nu_N)^m/m!\}.$$

Since (18) clearly implies that

$$\nu_n = O(\log n) \quad \text{as} \quad n \to +\infty,$$

it follows that, if p is fixed and $N \to +\infty$,

$$\sum_{n=1}^{N} | L(2\pi n) |^p = \sum_{m=0}^{\infty} \{\exp(-p/100) \cdot O(\log N)\}^m/m!\,.$$

Consequently, there exists for every $\epsilon > 0$ a sufficiently large integer $p = p_\epsilon > 0$ such that, as $N \to +\infty$,

$$\sum_{n=1}^{N} | L(2\pi n) |^p = O\{\sum_{m=0}^{\infty} (\epsilon \log N)^m/m!\,\} = O\{\exp(\epsilon \log N)\},$$

and so, by (8),

$$\sum_{n=-N}^{N} | L(2\pi n) |^p = 1 + 2O\{\exp(\epsilon \log N)\} = O(N^\epsilon).$$

This proves (16). And (17) is proved in the same way, if, starting with an arbitrary positive n which need not be an integer, one applies (18) to the integral part $[n]$ of n, instead of to n. Then (19) remains valid if the subscript n of μ_n is replaced by $[n]$; so that the balance of the above proof can be repeated without change.

7. All that remains to be done is what is assigned by (iii), § 2.

If $H = H(x)$, $-\infty < x < +\infty$, is a distribution function, (i. e., a monotone function such that $H(-\infty) = 0$, $H(+\infty) = 1$), let $\Lambda(u; H)$ denote its Fourier-Stieltjes transform,

$$\Lambda(u; H) = \int_{-\infty}^{+\infty} e^{iux} dH(x); \qquad -\infty < u < +\infty. \tag{20}$$

Then

$$\Lambda(u; H_{\mathrm{I}} * H_{\mathrm{II}}) = \Lambda(u; H_{\mathrm{I}})\Lambda(u; H_{\mathrm{II}}), \tag{21}$$

if $H_{\mathrm{I}} * H_{\mathrm{II}}$ denotes the convolution

$$\int_{-\infty}^{+\infty} H_{\mathrm{I}}(x - y)\, dH_{\mathrm{II}}(y), \tag{22}$$

which is a distribution function for any pair H_{I}, H_{II} of distribution functions. It is implied by (21) that

$$H_{\mathrm{I}} * H_{\mathrm{II}} = H_{\mathrm{II}} * H_{\mathrm{I}}, \qquad (H_{\mathrm{I}} * H_{\mathrm{II}}) * H_{\mathrm{III}} = H_{\mathrm{I}} * (H_{\mathrm{II}} * H_{\mathrm{III}}),$$

and that, for any positive integer p,

$$(23) \qquad \Lambda(u; H^{(p)}) = (\Lambda(u; H))^p,$$

if, starting with an H, one puts

$$(24) \qquad H^{(1)} = H, \quad H^{(2)} = H^{(1)} * H, \quad H^{(3)} = H^{(2)} * H, \cdots.$$

Let, in particular, $H(x)$ be the distribution function which is 0 if $x \leqq -\frac{1}{2}$ and 1 if $x \geqq \frac{1}{2}$, while if $|x| \leqq \frac{1}{2}$, then $H(x)$ is the classical Cantor function, obtained in the usual manner by successive trisections of the interval $-\frac{1}{2} \leqq x \leqq \frac{1}{2}$. Then $H(x)$ is, for $-\infty < x < +\infty$, the infinite convolution

$$H(x) = B(3x) * B(3^2 x) * \cdots * B(3^k x) * \cdots,$$

where $B(x)$ is the distribution function which has a jump of $\frac{1}{2}$ at $x = -1$ and $x = 1$; so that $\Lambda(u; B) = \cos u$. Correspondingly,

$$(25) \qquad \Lambda(u; H) = \prod_{k=1}^{\infty} \cos(3^{-k} u).$$

Furthermore, from (24),

$$(26) \qquad H^{(p)}(x) = B(a_{1p} x) * B(a_{2p} x) * \cdots * B(a_{kp} x) * \cdots$$

where

$$(27) \qquad a_{kp} = 3^j, \text{ if } (j-1)p < k \leqq jp, \text{ where } j = 1, 2, \cdots.$$

Also, from (23), (25) and (8),

$$(28) \qquad \Lambda(u; H^{(p)}) = (L(u))^p.$$

Since it is clear from (8) that $L(3^h \cdot 2\pi)$ is independent of h for $h = 1, 2, \cdots$, and that $L(6\pi) > 0$, it follows from (28) that

$$(29) \qquad \Lambda(3^h \cdot 2\pi; H^{(p)}) = (L(6\pi))^p > 0 \text{ for } h = 1, 2, \cdots.$$

Now, the distribution function $H(x)$ is, if $-\frac{1}{2} \leqq x \leqq \frac{1}{2}$, the Cantor function, so that $H(-\frac{1}{2}) = 0$, $H(\frac{1}{2}) = 1$, and so $H(x)$ is continuous for $-\infty < x < +\infty$. Hence it is clear from the definition (20) of a convolution that each of the iterated functions (24) is continuous for $-\infty < x < +\infty$. On the other hand, $H^{(p)}(x)$ is, in view of (26), an infinite convolution of step functions. Consequently (cf. Jensen and Wintner [3], Theorem 35), the distribution function $H^{(p)}(x)$ either is singular or absolutely continuous. Since (29) assures that $\Lambda(u; H^{(p)}) \to 0$, $u \to \pm\infty$, is impossible for a fixed p, it follows that $H^{(p)}(x)$ is singular for every p.

Furthermore, $H(x)$ is constant outside the interval $-\frac{1}{2} \leqq x \leqq \frac{1}{2}$. Hence it is clear from (27) and from the definition (22) of a convolution that $H^{(p)}(x)$ is constant outside the interval $-\frac{1}{2}p \leqq x \leqq \frac{1}{2}p$. Accordingly, by (20),

$$\Lambda(u; H^{(p)}) = \int_{-\frac{1}{2}p}^{\frac{1}{2}p} e^{iux} dH^{(p)}(x).$$

It follows, therefore, from (28) and (17) that there exists for every $\epsilon > 0$ a sufficiently large $p = p_\epsilon$ such that, as $N \rightarrow +\infty$,

$$\sum_{n=-N}^{N} \left| \int_{-\frac{1}{2}p}^{\frac{1}{2}p} e^{ianx} dH^{(p)}(x) \right| = O(N^\epsilon)$$

holds uniformly for $\pi \leqq \alpha \leqq 2\pi$.

Thus, on keeping ϵ fixed (hence also $p = p_\epsilon$), and then changing the unit along the x-axis in such a way that the interval $-\frac{1}{2}p \leqq x \leqq \frac{1}{2}p$ becomes the interval $-\pi \leqq x \leqq \pi$, one has

$$(30) \qquad \sum_{n=-N}^{N} \left| \int_{-\pi}^{\pi} e^{inx} dG(x) \right| = O(N^\epsilon),$$

if $G = G_\epsilon$ denotes the monotone function which results from $H^{(p)}$, where $p = p_\epsilon$, upon the change of units. Since $H^{(p)}$ is singular, so is G. Hence, comparison of (30) and (1) assures the existence of a singular monotone G which satisfies (7).

This completes the proof of (5), i. e., of $\kappa^* = \frac{1}{2}$.

8. While the Cantor function $H(x)$ is constant almost everywhere, (27) implies that the singular monotone function (26) is, if $p > 1$, nowhere constant on an interval (cf. Kershner and Wintner [4], p. 547). Consequently, the above result implies one recently proved by Hartman and Kershner [2].

Actually, the above constructions can easily be extended to the case in which the sequence of Fourier-Stieltjes coefficients is replaced by the corresponding Fourier-Stieltjes transform. To this end, all that is needed is to replace in (10) the summation by an integration, and then, with the help of (17), to repeat the above proof with obvious modification. Then it follows that there exists for every $\epsilon > 0$ a singular monotone function $F(x)$, $-\pi \leqq x \leqq \pi$, such that

$$(31) \qquad \int_{-\pi}^{\pi} e^{iux} dF(x) = O(|u|^{-\frac{1}{2}+\epsilon}) \quad \text{as} \quad u \rightarrow \pm\infty,$$

CALCUTTA CENTRAL LIBRARY

where u is a continuous variable. It is also seen that (31) can be chosen as an even function, i. e., such that the curve $y = F(x)$ is symmetric with respect to the line $y = F(0)$.

MASSACHUSETTS INSTITUTE OF TECHNOLOGY,
THE JOHNS HOPKINS UNIVERSITY.

REFERENCES.

[1] T. Carleman, *Sur les équations intégrales singulières à noyau reél et symétrique,* Uppsala, 1923.

[2] P. Hartman and R. Kershner, "On the Fourier-Stieltjes transform of a singular function," *American Journal of Mathematics,* vol. 60 (1938), pp. 459-462.

[3] B. Jessen and A. Wintner, "Distribution functions and the Riemann zeta function," *Transactions of the American Mathematical Society,* vol. 38 (1935), pp. 48-88.

[4] R. Kershner and A. Wintner, "On symmetric Bernoulli convolutions," *American Journal of Mathematics,* vol. 57 (1935), pp. 541-548.

[5] R. Kershner, "Determination of a van der Corput-Landau absolute constant," *ibid.,* vol. 57 (1935), pp. 840-846.

[6] R. Kershner, "On singular Fourier-Stieltjes transforms," *ibid.,* vol. 58 (1936), pp. 450-452.

[7] J. E. Littlewood, "On the Fourier coefficients of functions of bounded variation," *Quarterly Journal of Mathematics,* vol. 7 (1936), pp. 219-226.

[8] D. Menchoff, "Sur l'unicité du développement trigonométrique," *Comptes Rendus,* vol. 163 (1916), pp. 433-436.

[9] N. Wiener, *The Fourier Integral and Certain of Its Applications,* Cambridge, 1933, p. 153.

[10] A. Wintner, "On a class of Fourier transforms," *American Journal of Mathematics,* vol. 58 (1936), pp. 45-90.

ON THE CHARACTERIZATION OF A DISTRIBUTION FUNCTION BY ITS FOURIER TRANSFORM.*

By R. P. BOAS, JR.[1] and F. SMITHIES.

1. By a distribution function we shall mean a non-decreasing function $\sigma(x)$ defined on $-\infty < x < \infty$ such that $\sigma(-\infty) = 0$, $\sigma(\infty) = 1$. The value of $\sigma(x)$ at a point of discontinuity will be assumed to be fixed by the normalization $\sigma(x) = \frac{1}{2}[\sigma(x+0) + \sigma(x-0)]$.

If $\sigma(x)$ is a distribution function, its Fourier transform $\Lambda(u;\sigma)$ is defined by

$$\Lambda(u;\sigma) = \int_{-\infty}^{\infty} e^{iux} d\sigma(x) \qquad (-\infty < u < \infty). \tag{1.1}$$

If $\Lambda(u;\sigma)$ is given, $\sigma(x)$ is determined by the inversion formula[2]

$$\sigma(x) - \sigma(0) = \frac{1}{2\pi i} \int_{-\infty}^{\infty} \frac{1 - e^{-iux}}{u} \Lambda(u;\sigma)\, du, \tag{1.2}$$

where the integral is to be interpreted, if necessary, as a Cauchy principal value.

It is well known that there is a correlation between the smoothness of $\sigma(x)$ and the order of magnitude of $\Lambda(u;\sigma)$ as $|u| \to \infty$. For example, if

$$\Lambda(u;\sigma) = O(|u|^{-k}) \qquad (|u| \to \infty)$$

for every $k > 0$, $\sigma(x)$ is of class C^∞ in $(-\infty, \infty)$; again, if

$$\Lambda(u;\sigma) = O(e^{-c|u|}) \qquad (|u| \to \infty) \tag{1.3}$$

for some $c > 0$, $\sigma(x)$ is analytic on $(-\infty, \infty)$, and, furthermore, there is a function $\sigma(z)$ of the complex variable $z = x + iy$, such that $\sigma(z)$ is analytic and bounded in every strip $|y| \leqq c' < c$ and coincides with $\sigma(x)$ on the real axis.

It is known that if (1.3) is replaced by the weaker condition

$$\Lambda(u;\sigma) = O\{\exp(-c|u|^{\alpha})\} \qquad (|u| \to \infty)$$

for some $\alpha < 1$, $\sigma(x)$ is not necessarily analytic on the whole real axis; for

---

* Received April 11, 1938.

[1] National Research Fellow.

[2] See, e. g., S. Bochner, *Vorlesungen über Fouriersche Integrale*, 1932, p. 66.

example, the function $\exp(-c\,|\,u\,|^{\alpha})$ $(0 \leqq \alpha < 1)$ is the Fourier transform of a distribution function which fails to be analytic at $x = 0$.[3]

In this note we discuss by means of examples just how much $\sigma(x)$ is restricted by the behavior of $\Lambda(u;\sigma)$ as $|\,u\,| \to \infty$. Our results may be stated descriptively as follows. No condition weaker than (1.3) ensures that $\sigma(x)$ has an interval of analyticity. Let $\eta(u)$ be any function which, when $|\,u\,| \to \infty$, decreases to zero more slowly than every function $e^{-c|u|}$ $(c > 0)$. Then there exists a distribution function $\sigma(x)$ whose points of non-analyticity form an arbitrary closed set (which may in particular be the empty set or the whole real axis), and whose Fourier transform $\Lambda(u;\sigma)$ is of the same order of magnitude as $\eta(u)$ when $|\,u\,| \to \infty$. Furthermore, there is a distribution function, belonging to a general Denjoy-Carleman quasi-analytic class on $(-\infty, \infty)$, and analytic except on an arbitrary closed set.

It is not known whether (1.3) is a necessary condition for $\sigma(z)$ to be analytic and bounded in the strip $|\,y\,| < c$. It may therefore be of interest that if $\sigma(z)$ is analytic and bounded in a portion of the complex plane somewhat more extensive than a strip, we can obtain an estimate for $\Lambda(u;\sigma)$ of the same type as (1.3). We shall, in fact, show that if $\sigma(z)$ is analytic and bounded in the "double funnel" $R = R_1 + R_2 + R_3$, where

$$\left.\begin{aligned} R_1 &= E_z[\,|\,y\,| < c\,], \\ R_2 &= E_z[\,|\,y\,| < |\,x - a\,|^{\gamma},\ x \geqq a\,], \\ R_3 &= E_z[\,|\,y\,| < |\,x + a\,|^{\gamma},\ x \leqq -a\,], \end{aligned}\right\}$$

and $c > 0$, $\gamma > 0$, $a \geqq 0$, then

$$\Lambda(u;\sigma) = O(|\,u\,|^{3/2} e^{-c|u|}) \qquad (|\,u\,| \to \infty). \tag{1.4}$$

The function $\sigma(z) = \frac{1}{2} + \frac{1}{\pi} \arctan z$ is analytic except at $z = \pm i$, $\sigma(x)$ is a distribution function, and $\Lambda(u;\sigma) = e^{-|u|}$; consequently we cannot expect to obtain an estimate much stronger than (1.4).

2. THEOREM 1. *Let* $\epsilon(u)$ *be an even function of* u *such that* $\epsilon(u) > 0$ *for all* u, $\epsilon(u) \downarrow 0$ *as* $u \to \infty$, $u\epsilon(u) \uparrow \infty$ *as* $u \to \infty$ *in* $u \geqq u_0 > 0$, *and* $e^{-u\epsilon(u)}$ *is convex in* $u \geqq u_0$. *Then there exists a distribution function* $\tau(y)$, *such that* $y = 0$ *is a point of non-analyticity of* $\tau(y)$, *and*

$$\Lambda(u;\tau) = O(e^{-|u|\epsilon(u)}) \qquad (|\,u\,| \to \infty). \tag{2.1}$$

The result of Theorem 1 will be made more precise in § 4, but the construction used here possesses independent interest.

---

[3] A. Wintner, "On a class of Fourier transforms," *American Journal of Mathematics*, vol. 58 (1936), pp. 45-90, 65.

We define an even function $L(u)$ by the relation

$$L(u) = \begin{cases} e^{-u\epsilon(u)} & (u \geqq u_0), \\ A + Bu & (0 \leqq u \leqq u_0), \end{cases}$$

where the constants A and B are chosen in such a way that $L(u)$ is convex and decreasing in $u > 0$. Set

$$\sigma(y) = \frac{1}{\pi}\int_0^\infty \frac{\sin uy}{u} L(u)\,du \qquad (-\infty < y < \infty),$$

where $\int_0^\infty$ is to be interpreted, if necessary, as $\lim\limits_{T\to\infty}\int_0^T$. Then

$$\tau(y) = A^{-1}\sigma(y) + \tfrac{1}{2}$$

is a distribution function with the desired properties.

Clearly,

$$\Lambda(u;\tau) = A^{-1}L(|\,u\,|),$$

so that (2.1) is satisfied.

Since $L(u)$ is convex and $L(u) \downarrow 0$ as $u \to \infty$,

$$\sigma'(y) = \frac{1}{\pi}\int_0^\infty \cos uy\, L(u)\,du \geqq 0 \qquad (y \neq 0)^4$$

and $\sigma(y) > 0$ when $y > 0$. Hence $\sigma(y)$, and therefore $\tau(y)$, are non-decreasing.

We have

$$\begin{aligned}\sigma(y) &= \frac{A}{\pi}\int_0^{u_0}\frac{\sin uy}{u}\,du + \frac{B}{\pi}\int_0^{u_0}\sin uy\,du + \frac{1}{\pi}\int_{u_0}^\infty L(u)\sin uy\,du \\ &= I_1 + I_2 + I_3.\end{aligned}$$

When $y \to \infty$, $I_1 \to \frac{1}{2}A$; $I_2 \to 0$; and, by the second mean-value theorem,

$$|\,I_3\,| \leqq 2L(u_0)/(\pi y) \to 0.$$

Hence $\sigma(y) \to \frac{1}{2}A$ as $y \to \infty$; similarly, $\sigma(y) \to -\frac{1}{2}A$ as $y \to -\infty$. Consequently $\tau(y)$ is a distribution function.

Set

$$f(z) = \frac{1}{\pi}\int_0^\infty e^{-uz}L(u)\,du \qquad (z = x + iy).$$

The abscissa of convergence of this Laplace integral is clearly $x = 0$. Since $L(u)$ is a positive function, $f(z)$ has a singularity at $z = 0$.[5] It follows that

$$\sigma'(y) = \Re[f(iy)]$$

---

[4] E. C. Titchmarsh, *Introduction to the theory of Fourier integrals*, 1937, p. 170.

[5] E. Landau, "Über einen Satz von Tschebyschef," *Mathematische Annalen*, vol. 61 (1905), pp. 527-550, 548.

is not analytic at $y = 0$. For, if $\sigma'(y)$ is analytic at $y = 0$, then $\sigma'(y) = g(iy)$, where $g(z)$ is analytic in a neighborhood of $z = 0$; $f(z) - g(z)$ is purely imaginary on $x = 0$, and consequently, by the Schwarz reflection principle, has an analytic extension in a neighborhood of $z = 0$. Hence $f(z)$ cannot have a singularity at $z = 0$, a contradiction.

3. Lemma 1. *Let $\lambda(t)$ be a positive function defined for $t > 0$, such that $\lambda(t) \downarrow 0$ as $t \to \infty$. Let $\mu(u)$ be the inverse function of $t/\lambda(t)$. Let $\kappa(n)$ be a real-valued function defined for $n = 1, 2, \cdots$. Let F be the closure of the set of points $\{\kappa(n) + i\lambda(n)\}$, and let E be the intersection of F with the real axis. Then there exists a distribution function $\sigma(x)$ such that E is the set of points of non-analyticity of $\sigma(x)$ and such that*

$$\Lambda(u;\sigma) = O(e^{-(1-\epsilon)\mu(|u|)}) \qquad (|u| \to \infty) \tag{3.1}$$

for any $\epsilon > 0$.

Set

$$\tau(z) = \tfrac{1}{2} + \tfrac{1}{\pi}\operatorname{arc\,tan} z \qquad (z = x + iy).$$

We cut the z-plane along the imaginary axis from i to $i\infty$ and from $-i$ to $-i\infty$. We choose the one-valued branch of $\tau(z)$ which is regular in the cut plane and which is determined by the condition

$$\tau(x) \to 1 \qquad (x \to \infty).$$

Then $\tau(x)$ is a distribution function; and $\tau(z)$ is bounded on any bounded closed set in the cut plane. Set

$$\sigma(z) = (e-1)\sum_{n=1}^{\infty} e^{-n}\,\tau\left\{\frac{z-\kappa(n)}{\lambda(n)}\right\}. \tag{3.2}$$

We cut the z-plane along the lines $x = \kappa(n)$ from $\kappa(n) + i\lambda(n)$ to $\kappa(n) + i\infty$ and from $\kappa(n) - i\lambda(n)$ to $\kappa(n) - i\infty$, and denote the closure of the sum of these cuts by H. The series in (3.2) is uniformly convergent on any bounded set at a positive distance from the set H, and also uniformly convergent on the real axis. $\sigma(x)$ is a distribution function, and E is the set of points of non-analyticity of $\sigma(x)$.

We have

$$\Lambda(u;\tau) = e^{-|u|};$$

hence

$$\Lambda\left[u;\tau\left\{\frac{z-\kappa(n)}{\lambda(n)}\right\}\right] = e^{iu\kappa(n)-|u|\lambda(n)};$$

consequently

$$\Lambda(u;\sigma) = (e-1)\sum_{n=1}^{\infty} e^{-n+iu\kappa(n)-|u|\lambda(n)}. \tag{3.3}$$

It follows that

$$(e-1)^{-1}\,|\,\Lambda(u;\sigma)\,| \leqq \sum_{n=1}^{\infty} e^{-n-|u|\lambda(n)} = \sum_{n=1}^{v} + \sum_{n=v+1}^{\infty}$$
$$\leqq ve^{-|u|\lambda(v)} + \sum_{n=v+1}^{\infty} e^{-n} \qquad (v=1,2,\cdots).$$

Since $t/\lambda(t)$ increases, we have $|\,u\,|\,\lambda(v) \geqq v$ whenever $v \leqq \mu(|\,u\,|)$; we take $v=[\mu(|\,u\,|)]$. Then

$$(e-1)^{-1}\,|\,\Lambda(u;\sigma)\,| \leqq ve^{-v}+e^{-v}$$
$$= O(e^{-(1-\epsilon)\mu(|u|)}) \qquad (|\,u\,| \to \infty)$$

for every $\epsilon > 0$.

LEMMA 2. *Let $\lambda(t)$ satisfy the hypotheses of Lemma 1. If $\kappa(n)=b\xi(n)$ $(n=1,2,\cdots)$ where b is a constant and $\xi(n)$ is an integer-valued function, and $\sigma(z)$ is defined by (3.2), then*

$$\limsup_{|u|\to\infty}\,|\,\Lambda(u;\sigma)\,|\,e^{2\mu(|u|)} > 0. \tag{3.4}$$

By (3.3), if $bu=2\pi h$, where h is an integer, we have

$$(e-1)^{-1}\Lambda(u;\sigma) = \sum_{n=1}^{\infty} e^{-n-|u|\lambda(n)} \geqq \sum_{n=v+1}^{\infty} e^{-n-|u|\lambda(n)},$$

where $v=[\mu(|\,u\,|)]$. When $n \geqq v+1$, $|\,u\,|\,\lambda(n) \leqq n$; hence

$$(e-1)^{-1}\Lambda(u;\sigma) \geqq \sum_{n=v+1}^{\infty} e^{-2n} = (e^2-1)^{-1}e^{-2[\mu(|u|)]}. \tag{3.5}$$

The result follows.

4. THEOREM 2. *Let $\delta(t)$ be a positive function defined for $t>0$, such that $\delta(t)\downarrow 0$ and $t\delta(t)\uparrow\infty$ as $t\to\infty$. Then there exists a distribution function $\sigma(x)$, analytic on $-\infty<x<\infty$, such that*

$$\limsup_{|u|\to\infty}\,|\,\Lambda(u;\sigma)\,|\,e^{2|u|\delta(|u|)} > 0, \tag{4.1}$$

and

$$\limsup_{|u|\to\infty}\,|\,\Lambda(u;\sigma)\,|\,e^{(1-\epsilon)|u|\delta(|u|)} < \infty \tag{4.2}$$

for every $\epsilon > 0$.

We observe that $e^{t\delta(t)}$ can be made to approach infinity as slowly as we please when $t\to\infty$.

We set $\mu(u)=u\delta(u)$, and

$$\lambda(t) = \frac{t}{\mu^{-1}(t)} = \frac{\mu(\mu^{-1}(t))}{\mu^{-1}(t)} = \delta(\mu^{-1}(t)).$$

It is clear that $\lambda(t) \downarrow 0$ as $t \to \infty$. The functions $\lambda(t)$ and $\mu(u)$ satisfy the conditions of Lemma 1. We take $\kappa(n) = n$ $(n = 1, 2, \cdots)$. The distribution function $\sigma(x)$ defined by (3.2) has, by Lemmas 1 and 2, the required properties.

THEOREM 3. *If $\delta(t)$ satisfies the conditions of Theorem 2, there exists a distribution function $\sigma(x)$, having $x = 0$ as its only point of non-analyticity, and such that $\Lambda(u;\sigma)$ satisfies* (4.1), (4.2).

The proof is parallel to that of Theorem 2, except that we take $\kappa(n) = 0$ $(n = 1, 2, \cdots)$. We note that $\sigma(x)$ has an extension which is analytic in each of the half-planes $\Re(z) > 0$, $\Re(z) < 0$.

THEOREM 4. *Let E be an arbitrary closed point set on the real axis. If $\delta(t)$ satisfies the conditions of Theorem 2, there exists a distribution function $\sigma(x)$, having E as its set of points of non-analyticity, and such that $\Lambda(u;\sigma)$ satisfies* (4.1), (4.2).

Let $\{\beta_n\}$ $(n = 1, 2, \cdots)$ be a sequence of points dense in E; let $\{p_n\}$ be a sequence of positive numbers whose sum is unity. Write $\rho(x)$ for the distribution function constructed in Theorem 3. Set

$$\sigma(z) = \sum_{n=1}^{\infty} p_n \, \rho(z - \beta_n). \tag{4.3}$$

Then $\sigma(x)$ is a distribution function; it is analytic on the complement of E because the series in (4.3) is uniformly convergent on any bounded point set at a positive distance from the lines $x = \beta_n$ $(n = 1, 2, \cdots)$; every point of E is a limit point of singularities of $\sigma(z)$, and hence a point of non-analyticity of $\sigma(x)$.

The Fourier transform of $\sigma(x)$ is

$$\Lambda(u;\sigma) = \Lambda(u;\rho) \sum_{n=1}^{\infty} p_n e^{iu\beta_n} = \Lambda(u;\rho)\Omega(u),$$

where $\Omega(u)$ is a uniformly almost periodic function. Since $\Lambda(u;\rho)$ satisfies (4.2), and $\Omega(u)$ is bounded, $\Lambda(u;\sigma)$ satisfies (4.2). Since for $\rho(x)$ the constant b of Lemma 2 vanishes, $\Lambda(u;\rho)$ satisfies (3.5) for every u. Therefore, since $\Omega(u)$ is uniformly almost periodic, $\Lambda(u;\sigma)$ satisfies (4.1).

5. Let $\epsilon(u)$ be a positive non-increasing function of u, defined in $(0, \infty)$, such that $u^{\frac{1}{2}}\epsilon(u)$ is non-decreasing for all sufficiently large u, and

$$\int^{\infty} \frac{\epsilon(u)}{u} \, du$$

is divergent. If $f(x) \varepsilon C^{\infty}(a, b)$, and

$$|f^{(k+1)}(x)|^{1/k} \leqq Ak/\epsilon(k) \qquad (a \leqq x \leqq b;\ k = 0, 1, 2, \cdots),$$

where A is independent of k and x, we shall say that $f(x)$ belongs to the class $Q\{\epsilon(u); a, b\}$. The classes $Q\{\epsilon(u); a, b\}$ are rather general Denjoy-Carleman quasi-analytic classes; they include, in particular, all the Denjoy classes, for which (for sufficiently large u)

$$\epsilon(u) = \frac{1}{\log u}, \qquad \epsilon(u) = \frac{1}{\log u \log\log u}, \cdots.$$

The class $Q\{1; a, b\}$ is the class of functions analytic on $a \leqq x \leqq b$.

It follows from a result of Ingham's[3] that if $\sigma(x)$ is a distribution function, and

$$\Lambda(u; \sigma) = O(e^{-c|u|\epsilon(|u|)}) \qquad (|u| \to \infty),$$

then $\sigma(x) \varepsilon Q\{\epsilon(u); a, b\}$ for any finite interval (a, b).

THEOREM 5. *Suppose that* $\epsilon(u)$ *satisfies the conditions stated above, and that* $\epsilon(u) \downarrow 0$ *as* $u \to \infty$. *Let* E *be an arbitrary closed point set on the real axis. Then there exists a distribution function* $\sigma(x)$ *such that* $\sigma(x) \varepsilon Q\{\epsilon(u); a, b\}$ *for every finite interval* (a, b), *and* E *is the set of points of non-analyticity of* $\sigma(x)$.

This is an immediate consequence of Theorem 4.

6. THEOREM 6. *Let* $c > 0$, $\gamma > 0$, $a \geqq 0$. *Let* R *be the open region* $R_1 + R_2 + R_3$ *of the* z*-plane* $(z = x + iy)$, *where*

$$\left.\begin{aligned} R_1 &= E_z[|y| < c], \\ R_2 &= E_z[|y| < |x - a|^{\gamma},\ x \geqq a], \\ R_3 &= E_z[|y| < |x + a|^{\gamma},\ x \leqq -a]. \end{aligned}\right\}$$

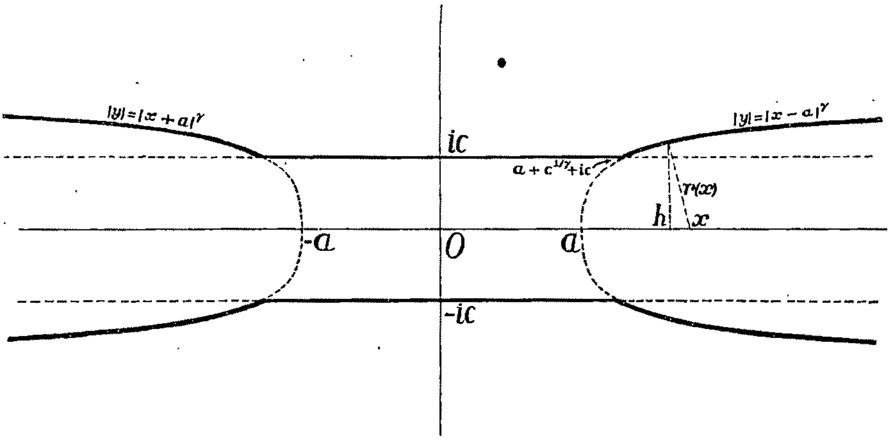
If $\sigma(z)$ *is analytic and bounded in* R, *and* $\sigma(x)$ *is a distribution function,*[7] *then*

$$\Lambda(u; \sigma) = O(|u|^{3/2} e^{-c|u|}) \qquad (|u| \to \infty). \tag{6.1}$$

---

[6] A. E. Ingham, "A note on Fourier transforms," *Journal of the London Mathematical Society*, vol. 9 (1934), pp. 29-32.

[7] We do not use the full force of the hypothesis that $\sigma(x)$ is a distribution function. The theorem remains valid if we assume instead that $\sigma'(x) \varepsilon L(-\infty, \infty)$, and define the function $\Lambda(u; \sigma)$ by (1.1).



Let $r(x)$ denote the distance from the point x of the real axis to the boundary of R. Then

$$|\sigma^{(n)}(x)| \leqq Mn!\,\{r(x)\}^{-n} \qquad (n=0,1,2,\cdots), \tag{6.2}$$

where M is a bound for $|\sigma(z)|$ in R. We have

$$r(x) \geqq c \qquad (-\infty < x < \infty). \tag{6.3}$$

Without loss of generality we may assume that $\gamma < \frac{1}{2}$. If $|x| > a$, it is obvious that, for some h such that $a < h < |x|$,

$$\{r(x)\}^{1/\gamma} = \{(|x|-h)^2 + (h-a)^{2\gamma}\}^{1/(2\gamma)}.$$

Hence, by Jensen's inequality,

$$\begin{aligned}\{r(x)\}^{1/\gamma} &\geqq (|x|-h)^{1/\gamma} + (h-a)\\ &= \lambda_1 + \lambda_2,\end{aligned}$$

say. When $|x|-h > 1$, $\lambda_1 > x-h$, so that $\lambda_1+\lambda_2 > |x|-h+h-a = |x|-a > |x|-a-1$. When $|x|-h \leqq 1$, we have $h \geqq |x|-1$, so that $\lambda_2 = h-a \geqq |x|-a-1$. Thus, in any case,

$$r(x) \geqq (|x|-a-1)^\gamma \qquad (|x| \geqq a+1). \tag{6.4}$$

From (1.1) we obtain

$$\Lambda(u;\sigma) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-iux}\sigma'(x)\,dx,$$

and hence, by integration by parts,

$$\Lambda(u;\sigma) = \frac{(-i)^k}{2\pi u^k}\int_{-\infty}^{\infty} e^{-iux}\sigma^{(k+1)}(x)\,dx \qquad (k=0,1,2,\cdots);$$

the integrated terms vanish by (6.2) and (6.4). Consequently

$$\begin{aligned}(6.5)\qquad 2\pi\,|\,u\,|^k\,|\,\Lambda(u;\sigma)\,| &\leqq \int_{-\infty}^{\infty}|\,\sigma^{(k+1)}(x)\,|\,dx\\ &=\int_{-\infty}^{-b}+\int_{-b}^{b}+\int_{b}^{\infty}\\ &=I_1+I_2+I_3,\end{aligned}$$

say, where $b=c^{1/\gamma}+a+1$. By (6.2) and (6.3),

$$(6.6)\qquad I_2\leqq 2M(k+1)\,!\,c^{-k-1}(c^{1/\gamma}+a+1)\qquad (k=0,1,2,\cdots).$$

By (6.2) and (6.4), for $k>1/\gamma$,

$$\begin{aligned}(6.7)\qquad I_1+I_3 &\leqq 2M(k+1)\,!\int_b^{\infty}(x-a-1)^{-\gamma(k+1)}dx\\ &=\frac{2M(k+1)\,!}{\gamma(k+1)-1}\,c^{-k-1+1/\gamma}.\end{aligned}$$

By (6.6) and (6.7),

$$I_1+I_2+I_3\leqq A(k+1)\,!\,c^{-k-1}\qquad (k>1/\gamma)$$

for some constant A. Hence by (6.5) and Stirling's formula,

$$(6.8)\qquad |\,\Lambda(u;\sigma)\,|\leqq A\,|\,u\,|\,(k+1)^{\frac{1}{2}}\left(\frac{k+1}{ce\,|\,u\,|}\right)^{k+1}\qquad (k>1/\gamma)$$

with a new A.

For any u such that $|\,u\,|>(2\gamma+1)/(\gamma c)$, we can take $k+1=[c\,|\,u\,|]$. It follows from (6.8) that

$$|\,\Lambda(u;\sigma)\,|\leqq A\,|\,u\,|^{3/2}e^{-c|u|}\qquad (|\,u\,|>(2\gamma+1)/(\gamma c)),$$

with a new A, i.e.

$$\Lambda(u;\sigma)=O(|\,u\,|^{3/2}e^{-c|u|})\qquad (|\,u\,|\to\infty).$$

This completes the proof.

PRINCETON UNIVERSITY AND
THE INSTITUTE FOR ADVANCED STUDY.

AN EXTENSION OF WIENER'S GENERAL TAUBERIAN THEOREM.*

By H. R. Pitt.

Suppose that $f(x)$ is defined in $(-\infty, \infty)$ and that two metrics $M\{f\}$, $M'\{f\}$ (which may be the same) satisfy the following postulates.

(a) $$M\{f\} = M\{|f|\}, \qquad M'\{f\} = M'\{|f|\}.$$

(b) $$0 \leq M\{f\} \leq M'\{f\}.$$

(c) $$M\{f_1 + f_2\} \leq M\{f_1\} + M\{f_2\}, \quad M'\{f_1 + f_2\} \leq M'\{f_1\} + M'\{f_2\}.$$

(d) $$M_x\{f(x-y)\} = M\{f\}, \quad M'_x\{f(x-y)\} = M'\{f\}.$$

(e) If $k(x)$ belongs to $L(-\infty, \infty)$ and $M'_x\{f(x, y)\} \leq C$ for all values of y, then $g(x) = \int k(y) f(x, y)\, dy$ exists for all values of x and

$$M\{g\} \leq \int |k(y)| \, M_x\{f(x,y)\} dy,$$
$$M'\{g\} \leq \int |k(y)| \, M'_x\{f(x,y)\} dy.$$

(Integrals in which limits are not specified are from $-\infty$ to ∞.)

Examples of pairs of metrics which satisfy these postulates are the following.

(A) $$M'\{f\} = \overline{bd}\, |f(x)|, \quad M\{f\} = \varlimsup_{x\to\infty} |f(x)|.$$

(B) $$M'\{f\} = M\{f\} = \overline{bd}\, |f(x)|.$$

(C) $$M'\{f\} = \overline{bd}\left[\int_x^{x+1} |f(y)|^p dy\right]^{1/p}, \quad M\{f\} = \varlimsup_{x\to\infty}\left[\int_x^{x+1} |f(y)|^p dy\right]^{1/p} \quad (p \geqq 1).$$

(D) $$M'\{f\} = \overline{bd}\left[\int_x^{x+1} |f(y)|^p dy\right]^{1/p}, \quad M\{f\} = \varlimsup_{x\to\infty}\left[\frac{1}{2x}\int_{-x}^{x} |f(y)|^p dy\right]^{1/p}.$$

(E) $$M'\{f\} = \overline{bd}\left[\int_x^{x+1} |f(y)|^p dy\right]^{1/p};$$
$$M\{f\} = \lim_{a\to\infty} \overline{bd}_{-\infty < x < \infty} \left[\frac{1}{a}\int_x^{x+a} |f(y)|^p dy\right]^{1/p}.$$

We shall prove the following theorem which includes Wiener's theorem when the metrics are defined as in (A).[1]

---

* Received May 17, 1938.

[1] H. R. Pitt, "General Tauberian theorems," *Proceedings of the London Mathematical Society*, vol. 43 (1938), Theorem 9, except that condition (4) is somewhat stronger than the condition $f(x) \epsilon T$.

THEOREM.

HYPOTHESIS. (1) $k(x)$ *belongs to* $L(-\infty, \infty)$,

(2) $K(x) = \int k(y) e^{-iyx} dy \neq 0 \qquad (-\infty < x < \infty)$;

(3) $M'\{f\} < \infty$,

(4) $M\{f(x+\epsilon) - f(x)\} = \delta(\epsilon), \qquad \lim_{\epsilon \to 0} \delta(\epsilon) = 0$,

(5) $g(x) = \int k(x-y) f(y) dy$.

CONCLUSION. *If* $C_1 > 0$, *we can choose* C_2, *depending only on* C_1, $k(x)$ *and* $\delta(\epsilon)$, *so that*

$$M\{f\} \leq C_1 + C_2 M\{g\}.$$

Let $\lambda > 0$,

$$h_\lambda(x) = \frac{\sin^2 \lambda x}{\pi \lambda x^2}.$$

Using a theorem[2] on Fourier transforms due essentially to Wiener, we can write

$$h_\lambda(y) = \int k(y-t) p_\lambda(t) dt,$$

where $p_\lambda(t)$ belongs to $L(-\infty, \infty)$. Now

$$G(x) = \int | k(x-y) f(y) | dy$$

exists for all values of x, by (1), (3) and postulates (d) and (e), and

$$M\{G\} \leq M\{f\} \int | k(x) | dx.$$

Hence

$$\int | p_\lambda(t) | dt \int | k(x-t) f(x-y) | dy = \int | p_\lambda(t) G(x-t) | dt$$

exists for all values of x, and it follows that

$$\begin{aligned} \int p_\lambda(t) g(x-t) dt &= \int p_\lambda(t) dt \int k(y-t) f(x-y) dy \\ &= \int f(x-y) dy \int k(y-t) p_\lambda(t) dt \\ &= \int h_\lambda(y) f(x-y) dy. \end{aligned}$$

If we write $P(\lambda) = \int | p_\lambda(y) | dy$, it follows from (d) and (e) that

$$(6) \qquad M\{\int h_\lambda(y) f(x-y) dy\} \leq M\{g\} \int | p_\lambda(y) | dy = M\{g\} P(\lambda).$$

---

[2] *Ibid.*, Theorem 6. When the paper was written I was not aware that Professor Wiener had given the same theorem in his course of lectures at Massachusetts Institute of Technology.

Since $\int h_\lambda(y)\,dy = 1$, we can write

$$M\{f\} = M\{\int h_\lambda(y) f(x-y)\,dy + \int h_\lambda(y)[f(x) - f(x-y)]\,dy\},$$

and using (4), (6) and (e), we have

$$\begin{aligned} M\{f\} &\leq P(\lambda)M\{g\} + \int h_\lambda(y) M\{f(x) - f(x-y)\}\,dy \\ &= P(\lambda)M\{g\} + \int h_\lambda(y)\delta(y)\,dy \\ &= P(\lambda)M\{g\} + \int h(y)\delta(y/\lambda)\,dy. \qquad (7) \end{aligned}$$

Since $\delta(y/\lambda) \leq 2M'\{f\}$, by (b) and (c), and $\lim_{\epsilon\to 0} \delta(\epsilon) = 0$, we can choose $\lambda = \lambda(C_1, \delta(\epsilon))$ so that

$$\int h(y)\delta(y/\lambda)\,dy \leq C_1,$$

and then (7) may be written

$$M\{f\} \leq C_1 + C_2 M\{g\},$$

where $C_2 = P(\lambda)$ depends only on C_1, $k(x)$ and $\delta(\epsilon)$.

HARVARD UNIVERSITY.

SYSTEMS OF DIFFERENTIAL EQUATIONS. I. THEORY OF IDEALS.*

By J. F. Ritt.

In publications of ourselves and of Raudenbush,[1] there was presented a theory of systems of differential equations which are algebraic in the unknowns and their derivatives. It is our object to start, with the present paper, the consideration of more general systems. We shall begin by treating systems of equations which are *analytic* in the independent variable and in the unknowns, and *algebraic* in the derivatives of the unknowns. Such a theory can be used as the basis for the study of any *finite* system of analytic differential equations, since, by replacing derivatives by new unknowns, the finite system can be made equivalent to a system involving only derivatives of the first order, and those algebraically.

This paper presents, for systems with the generality described above, a theory of ideals of the type given by Raudenbush for algebraic differential systems. In addition to the methods of our own earlier publications, we use the ideas in the very fundamental work of Raudenbush and also the methods of Rückert's important paper *Zum Eliminationsproblem der Potenzreihenideale.*[2] The present paper is complete in itself, except for the employment of Späth's extension of the Weierstrass preparation theorem. The use of Späth's theorem was suggested by Rückert's work. For unity of exposition, we have included a treatment of some questions which are covered by the work of Raudenbush. References to Raudenbush will indicate where we have borrowed from him.

Elements.

1. We shall study expressions in the unknown functions $y_1, \cdots, y_n$ of the independent variable x. We denote the j-th derivative of y_i by y_{ij}. We write, frequently, $y_i = y_{i0}$.

2. We consider any finite set of the y_{ij} with $j > 0$. Let a finite number

---

* Received February 21, 1938.

[1] Ritt, "Differential equations from the algebraic standpoint," *Colloquium Publications of the American Mathematical Society*, vol. 14 (New York, 1932); Raudenbush, "Ideal theory and algebraic differential equations," *Transactions of the American Mathematical Society*, vol. 36 (1934), pp. 361-368.

[2] *Mathematische Annalen*, vol. 107 (1933), p. 259.

of power products in these y_{ij} be formed and let each power product be multiplied by a power series in $x; y_1, \cdots, y_n$ which converges for numerically small values of those letters. The sum of the terms thus formed will be called an *element*. In other words, an element is a polynomial in any set of the derivatives (proper)[3] of the y_i, with coefficients which are functions of $x; y_1, \cdots, y_n$ analytic at $(0; 0, \cdots, 0)$.

Elements will be denoted by capital italic letters and systems of elements by large Greek letters.

We denote by Γ the totality of elements which are free of the independent variable x. Thus an element G is in Γ if G is a polynomial in certain y_{ij} with $j > 0$, with coefficients which are functions of $y_1, \cdots, y_n$ analytic at $(0, \cdots, 0)$.

Ordering of letters and elements.

3. We order the letters y_{ij} as follows. We shall call y_{ij} *higher* than y_{kl} if $j > l$ or if $j = l$ and $i > k$.

Let A be an element involving one more y_{ij} effectively. The highest letter y_{ij} in A will be called the *leader* of A.

4. We consider any element A and any definite y_{ij}. If A, written as a series of powers of y_{ij}, contains only a finite number of terms, we shall, as is natural, call A a *polynomial in* y_{ij}. Of course every element is a polynomial in every y_{ij} with $j > 0$.

5. Let A be a non-constant element in Γ (§ 2) and B any element in Γ.[4] If B is a constant, or if B is not a constant and has a leader which is lower than the leader of A, we shall say that A *is of higher rank than* B.

Now let A and B, in Γ and neither a constant, have the same leader y_{jk} and let A and B be polynomials in y_{jk}. If A is of higher degree than B in y_{jk}, we shall say, again, that A is of higher rank than B. If A and B have the same degree in y_{jk}, we shall say that A and B are of the same rank.

Any two constants will be considered to be of equal rank.

Thus, the only pairs A, B in Γ whose two elements we do not compare, are pairs in which A and B are not constants, and have the same leader, with at least one of A and B failing to be a polynomial in the common leader.

---

[3] Each y_i will be considered as its own derivative of order zero. Proper derivatives are derivatives of positive order.

[4] For our present purposes, only elements in Γ require ordering. The definition which follows immediately has somewhat more generality than is necessary for the applications.

Ascending sets.

6. Let A be a non-constant element in Γ which is a polynomial in its leader y_{jk}. An element B in Γ will be said to be *reduced with respect to* A if either

(a) $B = 0$

or

(b) *B is distinct from zero, contains no y_{jl} with*[5] $l > k$ *and is a poly-.*

7. A set of non-constant elements in Γ

$$A_1, \cdots, A_p \tag{1}$$

with $p > 1$ will be called an *ascending set* if every A_i is a polynomial in its leader and if, for $j > i$, A_j is higher than A_i and reduced with respect to A_i.

An element A_1 in Γ will be called an ascending set if A_1 is non-constant and is a polynomial in its leader, or if A_1 is a constant distinct from zero.

Let (1) be an ascending set with $p > 1$. Let A_i and A_j be two elements of (1) with $j > i$. Let y_{ab} be the leader of A_i and y_{cd} that of A_j. Because A_j is higher than A_i and is reduced with respect to A_i, y_{cd} must be higher than y_{ab}. It follows that a and c are distinct. Thus *the p leaders y_{ij} of the elements in* (1) *have distinct i.* This implies that $p \leqq n$.

The ascending set (1)[6] will be said to be *of higher rank* than the ascending set

$$B_1, \cdots, B_q$$

if either

(a) *there is a j, exceeding neither p nor q, such that A_i and B_i are of the same rank for $i < j$ and that A_j is higher than B_j,*[7] or

(b) *$q > p$ and A_i and B_i are of the same rank for $i \leqq p$.*

Two ascending sets for which no difference in rank is created by what precedes will be said to be of the same rank. For such sets $p = q$ and A_i and B_i are of the same rank for every i.

The above ordering of ascending sets is easily seen to be transitive.

8. We prove the following lemma:

---

[5] The j in y_{jl} is that which appears in y_{jk}.

nomial in y_{jk} with a degree in y_{jk} which is lower than that of A in y_{jk}.

[6] Here, as in all later work, it is understood that p may be unity.

[7] If $j = 1$, this is to mean that A_1 is higher than B_1.

LEMMA. *Every finite or infinite aggregate of ascending sets contains an ascending set whose rank is not higher than that of any other ascending set in the aggregate.*

Let us consider the first elements of all of the sets of the aggregate. We shall show that there is one of these first elements which is no higher than any other first element. This is certainly true if one of the first elements is a constant. Otherwise, we select those first elements which have a lowest leader and, from the elements just selected, we choose one which is of a lowest degree in the common leader.

Let us consider now those ascending sets in the given aggregate whose first elements have a least rank. Let σ_1 be the totality of such ascending sets. If the sets in σ_1 all consist of one element, any set in σ_1 will serve as the set whose existence was to be proved. Suppose that σ_1 contains sets which consist of more than one element. From among all such sets in σ_1 we select those whose second elements have a least rank and denote the totality of the sets selected by σ_2. If there are sets in σ_2 consisting of more than two elements, we make a third selection. As no ascending set has more than n elements, we arrive eventually at a collection of ascending sets of least rank.

9. Let (1) be an ascending set with A_1 non-constant. An element G in Γ will be said to be *reduced with respect to* (1) if G is reduced with respect to A_i, $i = 1, \cdots, p$.

Basic sets.

10. Let Σ be any system of elements in Γ which either contains a non-constant element which is a polynomial in its leader or else contains a constant distinct from zero. Because elements of the two types just mentioned are ascending sets, Σ contains ascending sets. By § 8, certain of the ascending sets in Σ have a least rank. Any such ascending set of least rank will be called a *basic set* of Σ.

Let Σ be a system in Γ for which (1), with A_1 non-constant, is a basic set. Then *no element in Σ which is an ascending set can be reduced with respect to* (1). Suppose that such an element, F, exists. Then F must be higher than A_1, else F would be an ascending set lower than (1). Similarly, F must be higher than A_2, else A_1, F would be an ascending set lower than (1). Finally, F is higher than A_p. Then $A_1, \cdots, A_p, F$ is an ascending set lower than (1). This proves our statement.

Let Σ be as in the preceding paragraph. We see that *if an element which is an ascending set and is reduced with respect to* (1), *is adjoined to* Σ, *the basic sets of the resulting system are lower than* (1).

Reduction.

11. Let G be any non-constant element in Γ which is a polynomial in its leader y_{ij}. We shall call the element $\partial G/\partial y_{ij}$ the *separant* of G. The coefficient of the highest power of y_{ij}, in G considered as a polynomial in y_{ij}, will be called the *initial* of G. G is higher than its separant and higher than its initial.

12. Let (1) be any ascending set with A_1 non-constant. Let S_i and I_i denote the separant and initial respectively of A_i, $i = 1, \cdots, p$.

We shall prove the following result.

Let G be an element in Γ which is a polynomial in the leader of every A_i and in all derivatives of each such leader. There exist non-negative integers s_i, t_i, $i = 1, \cdots, p$, which are such that when a suitable linear combination of the A_i and of derivatives,[8] *of certain orders, of the A_i, with elements for coefficients, is subtracted from*

$$S_1^{s_1} \cdots S_p^{s_p} I_1^{t_1} \cdots I_p^{t_p} G,$$

the remainder, R, is reduced with respect to (1).

Let u_i represent the leader of A_i in (1), $i = 1, \cdots, p$. We limit ourselves, as we may, to the case in which G involves derivatives, proper or improper, of one or more u_i. Let the highest letter in G which is a derivative of a u_i be represented by v and let v be a derivative of u_j. There is, of course, only one possibility for j.

To fix our ideas, we assume v higher than u_p. Using the algorithm of division, we find a relation

$$S_j^q G = CA'_j + B$$

where A'_j is a derivative of A_j, of some order, with v for leader; and where B is free of v. Because A'_j and S_j involve no letter higher than v, B involves no u_i-derivative[9] which is as high as v. For uniqueness of procedure, we take q as small as possible. We note that B is a polynomial in the u_i-derivatives.[10]

If B involves a u_i-derivative which is higher than u_p, we give B the treatment accorded to G. After a finite number of steps, we arrive at a unique element D which differs by a linear combination of derivatives of the A_i from an element

---

[8] By the j-th derivative of an element A, we mean the element which is obtained by differentiating A j times with respect to x, the y_i being regarded as functions of x.

[9] Thus we denote a derivative (proper or improper) of a u_i.

[10] If w is a letter in A'_j with respect to which A'_j is not a polynomial, then A_j cannot be a polynomial in w. Thus w is not a u_i-derivative.

$$S_1^{g_1} \cdots S_p^{g_p} G$$

and involves no u_i-derivative higher than u_p. D is a polynomial in the u_i-derivatives.

We then find a relation

$$I_p^{t_p} D = H A_p + K$$

where K is reduced with respect to A_p. K may involve u_p. Aside from u_p, the u_i-derivatives present in K are derivatives of $u_1, \cdots, u_{p-1}$. Such derivatives present in K are lower than u_p. Let v_1 be the highest of them.

Suppose that v_1 is higher than u_{p-1}. We give K the treatment accorded to G. In a finite number of steps we arrive at a unique element L which differs from some

$$S_1^{r_1} \cdots S_{p-1}^{r_{p-1}} I_{p-1}^{t_{p-1}} K$$

by a linear combination of A_{p-1} and proper derivatives of $A_1, \cdots, A_{p-1}$. L is reduced with respect to A_{p-1}.

Let us show that L is also reduced with respect to A_p. The first operation performed on K gives a relation

$$S_j^h K = E A'_j + K_1. \tag{2}$$

The highest letter in A'_j is v_1 and v_1 is lower than u_p. The highest letter in S_j is not higher than u_{p-1}. Thus, K_1 cannot contain a proper derivative of u_p. Let the degrees of K and K_1 in u_p be d and d_1 respectively. Suppose that $d_1 > d$. As S_j and A'_j are free of u_p, E in (2) must be of degree d_1 in u_p. Then EA'_j contains a term involving v_1 which is of degree d_1 in u_p. Such a term can be balanced neither by K_1 nor by the first member of (2). Thus K_1 is reduced with respect to A_p. Similar observations apply to the later steps in the calculation of L, so that L is reduced with respect to A_p.

Continuing, we reach a unique R as described above. We call R the *remainder of G with respect to* (1).

Series.

13. An element in Γ will be called a *series* if the element involves no y_{ij} with $j > 0$. A series is simply a function of $y_1, \cdots, y_n$ analytic at $(0, \cdots, 0)$.

Let A be a series which is not a constant. By the *class* of A, we shall mean the greatest j such that A involves y_j effectively. Constants will be called series of class zero.

Let A be a series of class $j > 0$. We shall call A *regular* if A contains a term ay_j^m with $m > 0$ and a a constant distinct from zero.

Regular systems.

14. Let Σ be a system of elements in Γ. We shall call Σ *regular* if either

(a) *A positive integer $j \leqq n$ exists such that Σ contains regular series of each of the classes $n, n-1, \cdots, j$ but contains no series of positive class less than j.*[11]

or

(b) *Σ contains no series of positive class.*

If (a) holds, we shall call $j-1$ the *index* of Σ. If (b) holds, the index of Σ will be defined as n.

15. Let Σ be any system of elements in Γ. We shall show that, by a non-singular linear transformation on $y_1, \cdots, y_n$, with constant coefficients, Σ can be converted into a regular system.[12]

If Σ contains no series of positive class, Σ is regular. Suppose that series of positive class exist in Σ and let A be one of them. By a non-singular linear transformation of the type

$$y_i = a_{i1}y'_1 + \cdots + a_{in}y'_n, \qquad (i = 1, \cdots, n),$$

with rational a_{ij}, we can convert A into a series B in the y'_i which is of class n and regular.[13] Under this transformation, Σ goes over into a system Σ' in the y'_i. If Σ' contains no series of one of the classes $1, \cdots, n-1$, Σ' is regular. Let Σ' contain a series C of positive class less than n. By a non-singular linear transformation applied to $y'_1, \cdots, y'_{n-1}$, we convert C into a regular series of class $n-1$. This second transformation converts B into a regular series of class n.

A sufficient number of repetitions of the above procedure furnishes a non-singular transformation which converts Σ into a regular system.

---

[11] If $j = 1$, this is to mean that Σ contains regular series of the classes $n, n-1, \cdots, 1$.

[12] Cf. Rückert, *loc. cit.*, p. 266.

[13] It is only necessary to regularize, in the manner explained in treatises of algebra, any one of the homogeneous polynomials of which A is the sum.

Residues.

16. The following important theorem is due to Späth.[14] *Let A and B be power series in $u_1, \cdots, u_p$ with numerical coefficients, convergent for $|u_i| < r$, $i = 1, \cdots, p$, where $r > 0$. For each j, let α_j, a series in $u_1, \cdots, u_{p-1}$, be the coefficient of $u_p{}^j$ in A arranged as a power series in u_p. Let a positive integer q exist such that $\alpha_0, \alpha_1, \cdots, \alpha_{q-1}$ vanish for $u_1 = \cdots = u_{p-1} = 0$, while α_q does not. There exists one and only one relation*

$$B = CA + D \tag{3}$$

where C and D are power series in $u_1, \cdots, u_p$ convergent for numerically small values of the u_i, and where no exponent of u_p in D is greater than $q - 1$.

That is,

$$D = \beta_0 u_p{}^{q-1} + \cdots + \beta_{q-1}$$

with the β_i power series in $u_1, \cdots, u_{p-1}$.

17. Let now Σ be a regular system of index $s < n$ (§ 14). Let

$$B_{s+1}, B_{s+2}, \cdots, B_n \tag{4}$$

be series in Σ with B_i of class i and regular, $i = s + 1, \cdots, n$.

Let G be any element in Σ. If B_n does not vanish for $y_1 = \cdots = y_n = 0$, we have $G = M_n B_n$ with M_n an element. If the vanishing does occur, we arrange G as a polynomial in the y_{ij} with $j > 0$ and apply Späth's theorem to the coefficients in G. We find a relation

$$G = M_n B_n + G_1 \tag{5}$$

with G_1 a polynomial in y_n as well as in the y_{ij} with $j > 0$. Considering G_1 as such a polynomial, we give it, with respect to B_{n-1}, the treatment accorded to G with respect to B_n. Continuing, we reach a relation

$$G = M_n B_n + \cdots + M_{s+1} B_{s+1} + H$$

with the M elements and H an element which is a polynomial in y_i, $s + 1 \leq i \leq n$. We shall call H the *residue of G with respect to* (4).

Bases.

18. Let Σ be any system of elements[15] and Φ a finite subset of Σ. We shall call Φ a *basis* of Σ if, for every element G in Σ, a positive integer p depending on G exists such that G^p is a linear combination of the elements in Φ and of their derivatives (of arbitrary orders), with elements for coefficients.

---

[14] *Journal für die reine und angewandte Mathematik*, vol. 161 (1929), p. 95.

[15] Not necessarily in Γ.

§§ 19-23 are devoted to the proof of the lemma:

LEMMA. *Every infinite system of elements has a basis.*

We shall prove this lemma first for systems of elements in Γ. It will then be easy to pass to the general case. We note that if a system Σ of elements in Γ has a basis Φ, the expression of any G^p (as above) in terms of the elements of Φ may be assumed to have coefficients in Γ. In short, one may replace x, in the coefficients, by 0.

19. The notation

$$P \equiv Q \qquad (C_1, \cdots, C_p),$$

where all letters represent elements, will mean that $P - Q$ is a linear combination of $C_1, \cdots, C_p$ and their derivatives (of arbitrary orders), with elements for coefficients.[16]

The following lemma is due to Raudenbush:[17]

LEMMA. *If*

(6) $$(PQ)^g \equiv 0 \qquad (C_1, \cdots, C_p)$$

where g is some positive integer, a positive integer h exists such that

(7) $$(P'Q)^h \equiv 0 \qquad (C_1, \cdots, C_p),$$

where P' is the derivative of P.

The relation $P^gQ^g \equiv 0$ gives,[18] by differentiation,

$$gP^{g-1}P'Q^g + gP^gQ^{g-1}Q' \equiv 0.$$

Multiplying through by Q and dividing by g, we have

(8) $$P^{g-1}P'Q^{g+1} \equiv 0.$$

If $g = 1$, we have $(P'Q)^2 \equiv 0$. Suppose that $g > 1$. We differentiate (8), multiply through by $P'Q$ and use (6) and (8). We find that

$$P^{g-2}P'^3Q^{g+2} \equiv 0.$$

If $g = 2$ we have (7) with $h = 4$. In general, (7) holds with $h = 2g$.

---

[16] When the letters above represent elements in Γ, the coefficients can be taken in Γ.

[17] *Loc. cit.*, §§ 1, 2.

[18] All congruences are with respect to $C_1, \cdots, C_p$.

20. We now assume that there exist infinite systems of elements in Γ which have no bases. We work towards a contradiction.

LEMMA. *Let* Σ *be a system of elements which has no basis. Let* $F_1, \cdots, F_p$ *be elements such that when each element in* Σ *is multiplied by a suitable product of non-negative powers of* $F_1, \cdots, F_p$, *one obtains a system* Λ *which has a basis. Then the system*[19] $\Sigma + F_1F_2 \cdots F_p$ *has no basis.*[20]

Let $\Sigma + F_1 \cdots F_p$ have a basis. If, to a basis of a system, one adds any finite number of elements of the system, one secures a new basis for the system. Then let $\Sigma + F_1 \cdots F_p$ have a basis.

$$F_1 \cdots F_p;\, H_1, \cdots, H_r, \tag{9}$$

where the H_i are such that the elements of Λ which they yield, after the above described multiplications, form a basis for Λ. Let G be any element in Σ. Let K represent $F_1 \cdots F_p$. Because Λ has the basis mentioned above, we have, for some g,

$$(KG)^g \equiv 0 \qquad (H_1, \cdots, H_r). \tag{10}$$

We have, furthermore, for some h, referring to (9),

$$G^h = MK + M_1K' + \cdots + N_1H_1 + \cdots, \tag{11}$$

with K' the derivative of K, the unwritten terms involving either higher derivatives of K or a derivative, proper or improper, of an H_i.

Let t be a positive integer and let both sides of (11) be raised to the t-th power. If t is large, every term in the expression for G^{ht} which is free of the H_i and their derivatives will involve to a high power either K or some derivative of K appearing in (11). We multiply the expression for G^{ht} by a high power of G and then use (10) and § 19.[21] We see that a sufficiently high power of G is linear in the H_i and their derivatives. This means that $H_1, \cdots, H_r$ is a basis for Σ, so that the lemma is proved.

21. LEMMA. *Let* Σ *be a system of elements which has no basis. Let* $F_1, \cdots, F_p$ *be elements such that* $\Sigma + F_1 \cdots F_p$ *has a basis. Then at least one of the systems* $\Sigma + F_i$, $i = 1, \cdots, p$, *has no basis.*[20]

We may evidently limit ourselves to the case of $p = 2$. Let each of $\Sigma + F_1$, $\Sigma + F_2$ have a basis. Then we can find a set of elements $H_1, \cdots, H_r$

---

[19] The logical sum of Σ and $F_1 \cdots F_p$.

[20] Raudenbush, § 4.

[21] Actually, we are using a trivial extension of the result in § 19.

in Σ such that $F_1, H_1, \cdots, H_r$ is a basis for $\Sigma + F_1$ and $F_2, H_1, \cdots, H_r$ a basis for $\Sigma + F_2$. Let G be any element in Σ. Let g be such that G^g is linear in derivatives (proper or improper) of F_1 and the H_i and also linear in derivatives of F_2 and the H_i. If we multiply together these two expressions for G^g, we have G^{2g} expressed as a linear combination of certain $F_1{}^{(j)}F_2{}^{(k)}$, where superscripts denote differentiation, and of the H_i and their derivatives. Now, given any $F_1{}^{(j)}F_2{}^{(k)}$, a sufficiently high power of it is, by § 19, linear in F_1F_2 and its derivatives. We see now that a sufficiently high power of G^{2g} is linear in $F_1F_2, H_1, \cdots, H_r$ and their derivatives. Then $F_1F_2, H_1, \cdots, H_r$ is a basis for $\Sigma + F_1F_2$. This contradiction proves the lemma.

22. We now complete, for systems of elements in Γ, the proof of the lemma stated in § 18.

Let there exist a system in Γ which has no basis. By § 15, there will exist a regular system without a basis. Let s, with $0 \leqq s \leqq n$ be the least integer which is the index of a regular system without a basis.

We suppose first that $s < n$.

Let Σ be any regular system of index s which has no basis. We consider a set (4) contained in Σ. We observe that every B_i in (4) vanishes when its letters are replaced by zero. A B_i which did not so vanish would be a basis of Σ. By the Weierstrass preparation theorem, we have $B_i = A_iB'_i$, $i = s+1, \cdots, n$, with B'_i a regular series of class i which is a polynomial in y_i and A_i a series of class not exceeding i which does not vanish for zero values of its letters. When each B_i is replaced in Σ by B'_i, we obtain a regular system Σ_1. Σ_1 has no basis, for if it had a basis composed of the B'_i and other elements $H_1, \cdots, H_r$, Σ would have a basis composed of the B_i and of $H_1, \cdots, H_r$.

From each B'_j with $j > s+1$, we can subtract a linear combination of the B'_i with $i < j$, with series for coefficients, so as to obtain a regular series C_j of class j which is a polynomial in $y_{s+1}, \cdots, y_{j-1}$ as well as in y_j. Let Σ_2 result from Σ_1 by replacing each B'_j with $j > s+1$ by C_j. Then Σ_2 is a regular system of index s which has no basis. Let G be any element of Σ_2 which is not in

$$B'_{s+1}, C_{s+2}, \cdots, C_n. \tag{12}$$

Let H be the residue of G with respect to (12). Let Σ_3 be the system composed of (12) and of the totality of the H. Then Σ_3 has no basis. Σ_3 cannot contain a series of positive class less than $s+1$; if it did, it could be transformed into a regular system of index less than s, without a basis. Thus Σ_3 is regular and of index s.

There exist thus regular systems of index s which lack bases, in which each element of the system is a polynomial in every letter not lower than y_{s+1}. Of all such systems, let Σ be one whose basic sets are not higher than those of any other such system (§ 8). Let

$$(13) \qquad A_1, \cdots, A_p$$

be a basic set of Σ. Then A_1 is not a constant; if it were, it would be a basis of Σ. Let S_i and I_i be the separant and initial respectively of A_i, $i = 1, \cdots, p$. Let

$$(14) \qquad B_{s+1}, \cdots, B_n$$

be a set (4) in Σ.

Let G be any element of Σ which is not in (13) or (14). Let R be the remainder of G with respect to (13).

Let Ω be the system composed of (13), (14) and the totality of the R. We shall prove that Ω has a basis.

Let this be false. Then there are certainly non-zero R in Ω. If no R is a series of positive class less than $s + 1$, Ω is a regular system of index s, with each element a polynomial in every letter not lower than y_{s+1}, whose basic sets, by § 10, are lower than (13). This is incompatible with the absence of a basis. If some R is a series of positive class less than $s + 1$, we can transform Ω into a regular system of index less than s.

Thus Ω has a basis. Let Λ be the system composed of (13), (14) and of the products of the general type

$$(15) \qquad S_1^{s_1} \cdots I_p^{t_p} G$$

which are used in forming the R in Ω. Let a basis of Ω consist of (13), (14) and of certain R, say $R_1, \cdots, R_q$. Then Λ has a basis consisting of (13), (14) and the elements (15) which give $R_1, \cdots, R_q$.

The lemmas of §§ 20, 21 inform us now that at least one of the systems $\Sigma + S_i$, $\Sigma + I_i$ has no basis. If, in such a system without a basis, the S_i or I_i adjoined is not a series of positive class less than $s + 1$, the system is regular and has basic sets lower than (13). The case of a series of positive class less than $s + 1$ is handled by the usual transformation. The hypothesis that Σ has no basis is thus untenable in the case of $s < n$.

For $s = n$, we take Σ as a regular system, lacking a basis, whose basic sets are as low as possible. The proof proceeds as above, without the set (14).

23. Now let Σ be an infinite system of elements which are not necessarily in Γ. We shall prove that Σ has a basis. Let G be any element in Σ.

In the expression for G in terms of x and the y_{ij}, we replace x by a new letter y_0. Then G goes over into an element G' in the functions $y_0, y_1, \cdots, y_n$ of the variable x. Σ goes over, in this manner, into a system Σ' in $y_0, \cdots, y_n$, Σ' being of the type examined in the preceding sections. Let $H'_1, \cdots, H'_r$ be a basis of Σ', H'_i being derived, as above, from an element H_i in Σ.

Referring to any G and G' as above, let G'^t be linear in the H'_i and their derivatives. We consider the equation which expresses G'^t in terms of the H'_i. In this equation, we replace y_0 by x, y_{01} by 1 and y_{0j} with $j > 1$ by 0. There results an expression for G^t in terms of the H_i and their derivatives. Thus, $H_1, \cdots, H_r$ is a basis for Σ.

Ideals.

24. A system Σ is called an *ideal* if it has the following properties:

(a) *Given any finite set of elements of* Σ, *every linear combination of those elements, with elements for coefficients, is in* Σ.

(b) *Given any element in* Σ, *the derivative of the element is also in* Σ.

An ideal Σ is called *perfect*[22] if, whenever an element is such that some power of it belongs to Σ, the element belongs to Σ. An ideal Σ is called *prime* if whenever AB is in Σ, at least one of A and B is in Σ. Every prime ideal is perfect.

We prove the following theorem:[22]

Every perfect ideal of elements is the intersection of a finite number of prime ideals.

Let Σ be a perfect ideal. Suppose that Σ is not the intersection of a finite number of prime ideals. Then Σ is not prime. Let AA', but neither A nor A', belong to Σ.

Let $H_1, \cdots, H_r$ be any basis of Σ. Let G be any element such that, for some g,

$$G^g \equiv 0 \qquad (H_1, \cdots, H_r, A). \tag{16}$$

The details of the preceding sections show that the totality Σ_1 of such elements G is a perfect ideal. Using A', in place of A, we obtain similarly a perfect ideal Σ'_1.

We prove now that Σ is the intersection of Σ_1 and Σ'_1. We note first that Σ is a proper subset of each of Σ_1 and Σ'_1. Now let G be any element common to Σ_1 and Σ'_1. We consider (16), taking g large enough to have also

---

[22] Raudenbush, § 3.

$$(17) \qquad G^g \equiv 0 \qquad (H_1, \cdots, H_r, A').$$

As in § 21, we see that, for some t,

$$G^{2gt} \equiv 0 \qquad (H_1, \cdots, H_r, AA'),$$

which means that G is in Σ.

At least one of Σ_1 and Σ'_1 is not the intersection of a finite number of prime ideals. Let this be the case for Σ_1. We find, as above, a perfect ideal Σ_2 of which Σ_1 is a proper part. Continuing, we find an infinite sequence of perfect ideals,

$$(18) \qquad \Sigma, \Sigma_1, \Sigma_2, \cdots$$

each a proper part of its successor. Let Ω be the logical sum of the ideals in (18). Let $H_1, \cdots, H_r$ be a basis of Ω. There is some Σ_p which contains all of the H_i. That Σ_p contains Ω. This contradiction proves that Σ is the intersection of a finite number of prime ideals.

Now, let the perfect ideal Σ be the intersection of the prime ideals

$$(19) \qquad \Sigma_1, \cdots, \Sigma_r.$$

Suppressing some of the Σ_i if necessary, we suppose that no Σ_i contains any Σ_j with $j \neq i$. We shall then prove that (19) is *unique*, that is, that if Σ is the intersection of prime ideals

$$\Omega_1, \cdots, \Omega_s$$

with no Ω_i containing any Ω_j with $j \neq i$, then $r = s$ and each Σ_i is identical with some Ω_j.

We show first that Σ_1 contains some Ω_j. Let this be false. Let A_i be an element in Ω_i, $i = 1, \cdots, s$, which is not in Σ_1. Then $A_1 A_2 \cdots A_s$, which is in every Ω_i, hence in Σ, is not in Σ_1. This is impossible. Let, then, Σ_1 contain Ω_1. Now, Ω_1, similarly, contains some Σ_j which must be Σ_1 since Σ_1 cannot contain a Σ_j with $j \neq 1$. The uniqueness is proved.

COLUMBIA UNIVERSITY.

DETERMINATION OF A VAN DER CORPUT ABSOLUTE CONSTANT.*

By RICHARD KERSHNER.

A very useful lemma of van der Corput[1] is to the effect that there exist absolute constants μ, ν with the property that if $f(x)$ be a real valued function possessing, in the finite interval $[a, b]$, a second derivative nowhere less than a fixed positive constant r, then

$$\left| \int_a^b \exp(if(x))\,dx \right| \leqq \mu r^{-\frac{1}{2}} \tag{1}$$

and

$$\left| \int_a^b \cos f(x)\,dx \right| \leqq \nu r^{-\frac{1}{2}} \tag{2}$$

it being understood that μ, ν are independent of $[a, b]$ as well as of $f(x)$.

The author has given[2] a very elementary proof of this lemma in a way that yields the best possible value ν_0 of ν satisfying (2). In fact it was shown that the expression on the left of (2) attains a maximum, under the restriction $f''(x) \geqq r > 0$, if the function $f(x)$ is chosen to be a parabola $f(x) = rx^2/2 + c_0$, for a certain constant c_0, and if $-a = b = [(\pi - 2c_0)/r]^{\frac{1}{2}}$. The constant c_0 is determined as the only root c of the equation

$$\int_0^{(\pi/2-c)^{\frac{1}{2}}} \sin(x^2 + c)\,dx = 0 \tag{3}$$

in $-\pi/2 \leqq c \leqq \pi/2$. This explicit determination of the maximum of the left side of (2) furnished the decimal approximation $\nu_0 = 3.327 \cdots$ while the best estimate previously given seems to have been $\nu_0 \leqq 32^{\frac{1}{2}} = 5.657 \cdots$.

The object of this note is to determine the best constant $\mu = \mu_0$ satisfying the original van der Corput inequality (1). First it will be shown that the two problems are actually equivalent and in fact $\mu_0 = \nu_0$. Furthermore, a maximum for the expression on the left of (1), under the restriction $f''(x) \geqq r$, is taken if the same function $f(x)$ and interval $[a, b]$ are chosen as in the case

---

* Received February 10, 1938.

[1] Cf. E. Landau, *Vorlesungen über Zahlentheorie*, vol. 2 (1927), p. 60; or E. Landau, *Einfuhrung in die Differentialrechnung und Integralrechnung* (1934), p. 307.

[2] R. Kershner, "Determination of a van der Corput-Landau absolute constant," *American Journal of Mathematics*, vol. 57 (1935), pp. 840-846.

(2). Secondly, the determination of the best constant in (1) will be carried out directly, without reference to the equivalent problem connected with (2). In view of this equivalence, the direct treatment of the case (1) to be given here may be considered as an alternative to the author's previous treatment of (2). In this connection it may be mentioned that, while the considerations to be used are certainly no simpler than those previously used in connection with (2) they do have the advantage of being considerably more intuitive geometrically. Furthermore, the method to be used in the direct treatment of (1) seems to be of interest in itself and is expected to have further applications in the study of exponential integrals.

THEOREM I. *Let* μ_0, ν_0 *denote the least values of* μ, ν *satisfying* (1), (2) *respectively, for all functions* $f(x)$ *with* $f''(x) \geqq r > 0$. *Then* $\mu_0 = \nu_0$.

The trivial inequality

$$\left| \int_a^b \exp(if(x))\,dx \right| \geqq \left| \int_a^b \cos f(x)\,dx \right| \tag{4}$$

shows that $\mu_0 \geqq \nu_0$ so it is only necessary to show that $\mu_0 \leqq \nu_0$. Let $f(x)$ satisfy the conditions of the van der Corput lemma. Then

$$\int_a^b \exp(if(x))\,dx = \left| \int_a^b \exp(if(x))\,dx \right| \exp(i\omega)$$

where $\omega = \arg\left[\int_a^b \exp(if(x))\,dx\right]$ is a constant. Thus

$$\int_a^b \exp i(f(x) - \omega)\,dx = \left| \int_a^b \exp(if(x))\,dx \right|. \tag{5}$$

This shows that the left-hand side of (5) is real, and in fact positive, so that

$$\int_a^b \exp[i(f(x) - \omega)]\,dx = \left| \int_a^b \cos(f(x) - \omega)\,dx \right|. \tag{6}$$

From (5) and (6)

$$\left| \int_a^b \exp(if(x))\,dx \right| = \left| \int_a^b \cos(f(x) - \omega)\,dx \right| \tag{7}$$

so that $\mu_0 \leqq \nu_0$.

Comparison of the equality (7) with the above mentioned result concerning (2) makes the following statement obvious:

THEOREM II. *The expression on the left of* (1) *attains a maximum,* *under the restriction* $f''(x) \geqq r > 0$, *if the function* $f(x)$ *is chosen to be the parabola* $f(x) = rx^2/2$, *and if* $-a = b = [(\pi - 2c_0)/r]^{1/2}$.

Theorem II will now be demonstrated by a direct geometrical argument. Consider the curve $S(f)$, represented in the complex plane by

$$z = x + iy = \int_0^s \exp(if(x))\,dx$$

or with parametric equations

$$(8) \qquad \begin{aligned} x &= \int_0^s \cos f(x)\,dx \\ y &= \int_0^s \sin f(x)\,dx. \end{aligned}$$

CALCUTTA UNIVERSITY CENTRAL LIBRARY

Since $(dx/ds)^2 + (dy/ds)^2 = 1$, the parameter s is actually the arc-length along the curve $S(f)$. Also

$$dy/dx = \tan f(s)$$

so that $f(s)$ represents the inclination of $S(f)$ at the point s (i. e., at the point with arc-length s) and $f'(s)$ is the curvature of $S(f)$ at the same point.

It will be assumed that $f'(s)$ takes on all values between $-\infty$ and $+\infty$, since otherwise the region of definition of the function $f(x)$ can be extended. Also it will be assumed that $f(0) = f'(0) = 0$, a normalization that involves no loss of generality. Thus the curve $S(f)$ passes through the origin when $s = 0$ and is tangent to the x-axis at this point. Furthermore the curvature $k(s) = f'(s)$ of $S(f)$ vanishes at this point. For $s < 0$ the curvature $k(s)$ is negative and for $s > 0$, $k(s) > 0$. Finally $k'(s) = f''(s) \geqq r > 0$ by assumption.

It is easy to see that the curve $S(f)$ is a double spiral without double points (see figure p. 552). In fact, if the osculating circle be drawn at the point $s = a > 0$ then all points of the curve $S(f)$ for which $s > a$ lie within this circle while all points $s < a$ lie outside this circle. If $a < 0$ the situation is reversed. Finally, at $a = 0$ there is no osculating circle but the two arcs of $S(f)$ corresponding to $s > 0$ and $s < 0$ lie on opposite sides of the tangent line, i. e., of the x-axis.

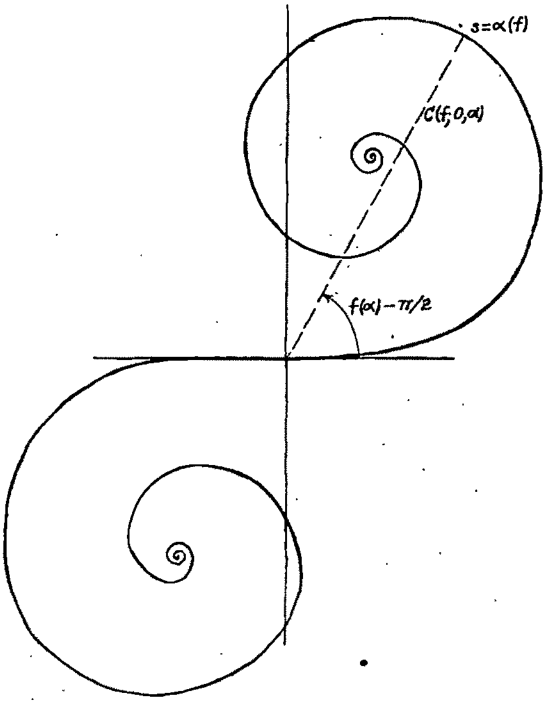
Now $\left| \int_a^b \exp(if(x))\,dx \right|$ is simply the length of the chord joining the points $s = a$, $s = b$ on the curve $S(f)$. The notation $C(f; a, b)$ will be used to denote this chord or its length. Thus the problem is to determine the maximum of $C(f; a, b)$ for all spirals $S(f)$.

First it will be shown that it is sufficient to consider chords which pass through the origin and spirals which are symmetrical with respect to the origin. Suppose that a and b are named in such a way that $C(f;0,a) \geqq C(f;0,b)$, $a \neq 0$. Then, from the triangle inequality

$$C(f;a,b) \leqq C(f;0,a) + C(f;0,b)$$

it is seen that

$$2C(f;0,a) \geqq C(f;a,b). \tag{9}$$



Let $g(x)$ be defined by $g(x) = f(x)$ if $\operatorname{sign}(x) = \operatorname{sign}(a)$, $g(-x) = g(x)$ and $g(0) = 0$. Then $S(g)$ is symmetrical with respect to the origin and there exists on $S(g)$ a chord $C(g;-a,a)$ such that

$$C(g;-a,a) = 2C(f;0,a) \geqq C(f;a,b).$$

Thus, to find a maximum for all chords $C(f;a,b)$ it is sufficient to find a maximum for $2C(f;0,a)$ where $a > 0$.

Now for a given $S(f)$ there does exist a maximum $C(f;0,a)$ and this maximum chord is, of course, perpendicular to the curve $S(f)$ at the point a. In fact, it is easily seen from the shape of $S(f)$ that the maximum $C(f;0,a)$ occurs when a is the smallest value for which $C(f;0,a)$ is perpendicular to the curve $S(f)$ at the point a. Let this value a be denoted by $a=\alpha=\alpha(f)$. The value $a=\alpha$ clearly occurs when the inclination $f(\alpha)$ is between $\pi/2$ and $3\pi/2$. Finally, the closed curve consisting of the arc of $S(f)$ for which $0\leqq s\leqq\alpha$ and the chord $C(f;0,\alpha)$ is a convex curve which will be denoted by $\Gamma(f;\alpha)$.

Now it will be shown that the maximum chord $C(f;0,\alpha)$ occurs when $f(x)=f_0(x)=rx^2/2$. It should be mentioned that $S(f_0)$ is the well known spiral of Cornu. Let β be such that $f_0(\beta)=f(\alpha)$. In conformity with the above let $\Gamma(f_0,\beta)$ denote the closed curve (also obviously convex) consisting of the arc of $S(f_0)$ for which $0\leqq s\leqq\beta$ and the chord $C(f_0;0,\beta)$.

Now $f''(x)\geqq r>0$ and $f''_0(x)\equiv r$. Also $f(0)=f_0(0)=0$; $f'(0)=f'_0(0)=0$. It has been shown,[3] under these conditions, that $f'(x_1)\geqq f'_0(x)$ if x_1 is the uniquely determined value $x_1=x_1(x)$ for which $f(x_1)=f_0(x)$. Interpreted geometrically this means that the curvature $k_0(s)$ of $S(f_0)$ at any point s is not greater than the curvature $k(s_1)$ of $S(f)$ at the point s_1 where the inclination $f(s_1)$ is the same as the inclination $f_0(s)$ of $S(f_0)$ at s.

Let ϕ denote the normal inclination of $\Gamma(f;\alpha)$ or $\Gamma(f_0;\beta)$ and let $K(\phi)$, $K_0(\phi)$ represent the curvature of these curves where the curvature exists, i. e., up to the ϕ-intervals corresponding to the corners, so that

$$K(\phi)=k(s(\phi)),\quad K_0(\phi)=k_0(s_0(\phi));\quad -\pi/2<\phi<f_0(\beta)-\pi/2.$$

Then, according to the preceding paragraph,

$$K(\phi)\geqq K_0(\phi);\quad -\pi/2<\phi<f_0(\beta)-\pi/2. \tag{10}$$

A well known theorem[4] on convex curves states that if the inequality occurring in (10) holds for all ϕ-values then the curve $\Gamma(f;\alpha)$ can be placed within $\Gamma(f_0;\beta)$. Slight modifications of the proof of this theorem, as given *loc. cit.*[4], show that if $H(\phi)$, $H_0(\phi)$ are the supporting functions[4] of $\Gamma(f;\alpha)$, $\Gamma(f_0;\beta)$ respectively then

$$H(f(\alpha)-\pi/2)=H(f_0(\beta)-\pi/2)\leqq H_0(f_0(\beta)-\pi/2). \tag{11}$$

But, by definition of the supporting function,

$$H(f(\alpha)-\pi/2)=C(f;0,\alpha)\quad\text{and}\quad H_0(f_0(\beta)-\pi/2)=C(f_0;0,\beta).$$

---

[3] *Loc. cit.*[2], p. 845.

[4] Cf. W. Blaschke, *Kreis und Kugel* (1916), pp. 115-116.

Thus, from (11),

$$C(f; 0, \alpha) \leqq C(f_0; 0, \beta). \tag{12}$$

It is now seen from (9) and (12) that

$$\begin{aligned} C(f; a, b) &\leqq 2 \max C(f; 0, a) = 2C(f; 0, \alpha) \\ &\leqq 2 \max C(f_0; 0, c) = \max C(f_0; -c, c). \end{aligned}$$

Expressed in the integral form, with the definition of $f_0(x)$ substituted, this last inequality becomes

$$\left| \int_a^b \exp(if(x))\, dx \right| \leqq \max \left| \int_{-c}^{c} \exp(irx^2/2)\, dx \right| .$$

The proof of Theorem II is completed by the trivial determination of c.

THE UNIVERSITY OF WISCONSIN.

A PROPERTY OF SPHERICAL HARMONICS.*

By Hans Lewy.

In this paper will be proved the following theorems:

Theorem 1. *Let* $u(x, y, z)$ *be a homogeneous polynomial of degree* $n > 2$ *whose range of values covers positive and negative numbers and whose Hessian* $H_{[u]}$ *vanishes only at* $x = y = z = 0$. *Then those points of the unit sphere* σ $(x^2 + y^2 + z^2 = 1)$ *for which* $u = 0$, *form two distinct closed convex curves*[1] *without singularities and diametral to each other.*

Theorem 2. *Let* $u(x, y, z)$ *be a spherical harmonic of degree* $n > 2$, *i. e. a regular solution of*

$$u_{xx} + u_{yy} + u_{zz} = 0$$

and of

$$xu_x + yu_y + zu_z = nu$$

which does not vanish identically. Then there exists a point distinct from the origin where the Hessian

$$H_{[u]} \equiv \begin{vmatrix} u_{xx} & u_{xy} & u_{xz} \\ u_{xy} & u_{yy} & u_{yz} \\ u_{xz} & u_{yz} & u_{zz} \end{vmatrix}$$

vanishes.

We employ the following

Lemma 1. *If* $v(x, y, z)$ *be a homogeneous polynomial of degree* $m > 1$ *and* (x_0, y_0, z_0) *an arbitrary point where* $v(x_0, y_0, z_0) \neq 0$, *then the Gaussian curvature of the surface* $v(x, y, z) = v(x_0, y_0, z_0)$ *has the same sign as* $vH_{[v]}$ *and vanishes where* $H_{[v]}$ *vanishes.*

Proof of the Lemma. $|v|^{1/m}$ is of degree 1 and analytic wherever $v \neq 0$. The first derivatives of $|v|^{1/m}$ are of degree 0. On writing down the homogeneity relations for these, we obtain three homogeneous equations whose determinant vanishes. This determinant is $H_{[|v|^{1/m}]}$ and we have for $v \gtrless 0$ resp.

---

* Received March 14, 1938.

[1] A closed curve without double points is called convex if either its "interior" or its "exterior" is convex, i. e. contains, together with two points P and Q, the smaller arc of the great circle through P and Q.

$$
\begin{aligned}
0 &= H_{[|v|^{1/m}]} \\
&= \pm \begin{vmatrix} v_{xx} + (1/m - 1)\dfrac{v_x^2}{v}, \cdots & & , \cdots \\ v_{xy} + (1/m - 1)\dfrac{v}{v_x v_y}, & v_{yy} + (1/m - 1)\dfrac{v_y^2}{v}, & \cdots \\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ , & \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ , & \cdots \end{vmatrix} \cdot \frac{|v|^{3(1/m-1)}}{m^3} \\
&= \pm \frac{|v|^{3(1/m-1)}}{m^3} \{H_{[v]} - (1/v)(1 - 1/m)[(v_{xx}v_{yy} - v_{xy}^2)v_z^2 + \cdots]\}.
\end{aligned}
$$

The expression $(v_{xx}v_{yy} - v_{xy}^2)v_z^2 + \cdots$ differs from the Gaussian curvature of $v = \text{const.}$ by a positive factor, and our Lemma follows.

Proof of Theorem I. Replacing, if necessary, u by $-u$, we may assume that $H_{[u]} > 0$ and that the curvature of the surfaces S_+ ($u = \text{const.} > 0$) is positive, the curvature of the surfaces S_- ($u = \text{const.} < 0$) is negative. Furthermore it is legitimate to suppose n even, since for n odd the Hessian is an odd function and the conditions of our Theorem cannot be satisfied. Since $H_{[u]} \neq 0$, the curves $u = 0$ on σ are without singularities and possess a non-vanishing geodetic curvature at each point (i. e. have no contact of higher order with a great circle). For it is a familiar fact of algebraic geometry that $H_{[u]} \neq 0$ implies that the projection of $u = 0$ from the origin O on the tangent plane T of σ at a point P of $u = 0$ has no singularity at P and its curvature at P does not vanish; for the coördinates (x, y, z) on σ can be interpreted as projective coördinates of the projection on T.

Consider that branch of $u = 0$ which passes through P. It must be a closed spherical curve since it does not possess any singularities, and its geodetic curvature does not vanish. Its "interior," therefore, is a convex domain B distinct from a hemisphere.[2]

We show $u > 0$ in B. Assume that a suitable rotation of the system of coördinates transforms the above point P into $(1, 0, 0)$ and the direction of $u = 0$ at P into $dx = dz = 0$. We conclude from the homogeneity of u that at P for $dx = dz = 0$, $dy \neq 0$

$$
\begin{aligned}
&0 = u = u_x, \quad xu_{xx} = u_{xx} = 0, \quad u_x dx + u_y dy + u_z dz = du = u_y dy = 0, \\
&u_y = 0, \quad xu_{xy} = u_{xy} = 0.
\end{aligned}
$$

---

[2] I owe to M. W. Fenchel the following information regarding this theorem about spherical curves of positive geodetic curvature: the theorem, though not explicitly stated in literature, follows immediately from i) a theorem by Möbius that a positively curved arc without singularities nor double points is contained in a hemisphere (Möbius, *Werke II*, pp. 183-187), and ii) the corresponding fact about planar domains bounded by convex curves to which it may be reduced by central projection.

By hypothesis

$$H_{[u]} = \begin{vmatrix} 0 & 0 & u_{xz} \\ 0 & u_{yy} & u_{yz} \\ u_{xz} & u_{yz} & u_{zz} \end{vmatrix} = -u_{xz}{}^2 u_{yy} > 0$$

whence

$$u_{yy} < 0, \quad u_{xz} \neq 0.$$

Now we have along $u = 0$ at P

$$dz = 0, \quad 0 = d^2u/dy^2 = u_{yy} + u_z d^2z/dy^2$$

where y is taken as independent variable. Thus $u_z d^2z/dy^2 > 0$, and u is positive on that side of a neighborhood of $u = 0$ against which $u = 0$ is concave and negative on the other side.

If $u = 0$ at P then $u = 0$ at the diametral point P'. But P and P' cannot lie on the same branch of $u = 0$. For choose in the neighborhood of $P\,(1, 0, 0)$ two points P_1 and P_2 on $u = 0$ with small $|\,y_{P_1}|$, $|\,y_{P_2}|$ and $y_{P_1} < 0$, $y_{P_2} > 0$. B is contained within a loon formed by the two great circles tangent to $u = 0$ at P_1 and P_2. The condition: $d^2z/dy^2 \neq 0$ on $u = 0$ in the neighborhood of P excludes that either circle passes through P. On the other hand P lies within the loon, and consequently P' lies outside of it and of B.[3]

Thus the number of disconnected branches of $u = 0$ is even and each branch bounds a convex domain in which $u > 0$. For suppose $u = 0$ inside of a branch C of $u = 0$, on a certain set of points E. Choose in E a point Q of minimum distance from C. Since $u > 0$ in the neighborhood of C for all points inside of C, Q has a positive distance from C. Thus there would exist a convex curve C', passing through Q and contained within C, in whose exterior necessarily $u < 0$ for a neighborhood of C'. This leads to a contradiction with the assumption that Q is a point of E of minimum distance from C.

Hence the domain $u < 0$ on σ consists of σ minus an even number of disjoint convex domains.

Consider again an arbitrary point of $u = 0$ in the above notation: $P(1, 0, 0)$, $u = u_x = u_{xx} = u_{xy} = u_y = 0$, $u_{xz} \neq 0$. Denote by $Q'(1, 0, z)$ a point for which $|\,z\,|$ is small and $u < 0$. The asymptotic directions of the surface $u =$ const. that passes through Q' are determined by

$$u_x dx + u_y dy + u_z dz = 0 \tag{1}$$

$$u_{xx} dx^2 + 2u_{xy} dx dy + u_{yy} dy^2 + 2u_{xz} dx dz + \cdots = 0. \tag{2}$$

---

[3] Also cf. Footnote 2.

The second relation expresses that $d^2x = d^2y = d^2z = 0$ imply $d^2u = 0$. On replacing dx by its value in (1) and substituting in (2) we obtain

$$\begin{aligned} 0 = A \equiv\ & u_x(u_{yy}dy^2 + 2u_{yz}dydz + u_{zz}dz^2) \\ & + (u_ydy + u_zdz)[(u_{xx}/u_x)(u_ydy + u_zdz) - 2u_{xy}dy - 2u_{xz}dz]. \end{aligned}$$

As Q' tends to P $(z \to 0)$, this equation reduces to $(dz/dy)^2 = 0$. For

$$\begin{aligned} \lim_{z\to 0} \frac{u_{xx}u_z - 2u_xu_{xz}}{u_x} &= \frac{u_{xxz}u_z + u_{xx}u_{zz} - 2u_{xzz}u_x - 2u_{xz}u_{xz}}{u_{xz}}\bigg|_{z=0,\,x=1,\,y=0} \\ &= \frac{u_z^2[(n-1)(n-2) - 2(n-1)^2]}{(n-1)u_z}\bigg|_{x=1,\,y=0,\,z=0} \\ &= -nu_z(1,0,0), \end{aligned}$$

$$\lim_{z\to 0} \frac{u_{xx}u_y - u_{xy}u_x}{u_x} = 0.$$

Hence the projections of the asymptotic directions at Q' on T tend to the direction of the tangent of $u = 0$ at P as $Q' \to P$. The same result would have been obtained if, instead of $Q'(1, 0, z)$, we had considered the point $\bar{Q}$ $(1/\sqrt{1 + z^2},\ 0,\ z/\sqrt{1 + z^2})$ which is the intersection of OQ' with σ. For the asymptotic directions are determined by the equations (1) and (2) which are homogeneous in x, y, z.

We next consider the projections from O on the sphere σ of the asymptotic directions at $\bar{Q} = \bar{Q}(\bar{x}, 0, \bar{z})$, given by

$$(d'x, d'y, d'z) = (dx, dy, dz) - (\bar{x}dx + \bar{z}dz)(\bar{x}, 0, \bar{z})$$

with (dx, dy, dz) satisfying (1) and (2). We have

$$\lim_{\bar{z}\to 0} u_x(\bar{x}, 0, \bar{z})/\bar{z} = u_{xz}(1, 0, 0),$$

whence

$$\bar{z}\frac{dx}{dy} = -\bar{z}\frac{u_ydy + u_zdz}{u_xdy} \to \lim_{\bar{z}\to 0} \frac{dz}{dy} \cdot \frac{1}{n-1}.$$

Since $\lim\limits_{\bar{z}\to 0} \dfrac{dz}{dy} = 0$, we conclude

$$\frac{d'z}{d'y} = \frac{dz - (\bar{x}dx + \bar{z}dz)\bar{z}}{dy} \to 0$$

and

$$\frac{d'x}{d'y} = \frac{dx - (\bar{x}dx + \bar{z}dz)\bar{x}}{dy} \to \lim_{\bar{z}\to 0} \frac{(1 - \bar{x}^2)dx}{dy} = \lim_{\bar{z}\to 0} \bar{z}^2\frac{dx}{dy} = 0.$$

Thus is shown that the projections on σ of the asymptotic directions at points $\bar{Q}$ of σ with $u_{\bar{Q}} < 0$ form a field of two vectors which tend to the tangent of C as $\bar{Q}$ approaches C on a normal. It is easily seen that the above limit considerations hold uniformly for all points of C as the higher derivatives of u

on σ are uniformly bounded. The fields of the two vectors is therefore defined and continuous in the domains $u \leqq 0$.

Choose an arbitrary point N of σ for which $u_N > 0$ and project the vector field stereographically from N on a plane. Of the images of the spherical curves $u = 0$, one will contain the image of the domains $u < 0$ while the images of the other curves $u = 0$ will constitute interior boundaries of the domain of definition of the projected field vectors. As the field is without singularities the sum of the rotations of the vectors over the interior boundaries equals that over the exterior boundary. Evidently the contribution of each boundary is 4π. Hence there can be only one interior boundary. In other words, the curves $u = 0$ on σ are precisely two diametrally opposite convex curves without singularities.

The remaining paragraphs will show that for no spherical harmonic u of degree $n > 2$, n even, the curves $u = 0$ on σ can be two distinct diametrally opposite convex curves without singularities.

LEMMA 2. *Consider on σ a closed convex curve C_1 different from a great circle. Denote by B_1 the interior of C_1, and by B_2 the interior of the curve C_2 diametrally opposite to C_1. Then it is possible to locate on σ a circular domain of spherical radius 33° in such a way that its closure lies either in B_1 or in $\sigma - (B_1 + C_1) - (B_2 + C_2)$.*

We omit a proof of this Lemma and refer instead to a note by R. M. Robinson[4] in which is shown that one always can locate in the described way a circular domain of radius $< \arctan \frac{1}{2}\sqrt{3}$ and that this is the best possible upper bound.

LEMMA 3. *Let u be a spherical harmonic of degree $n \geqq 4$. Then each circular domain β of the unit sphere of radius 33° contains a zero of u.*

Suppose there exists a spherical harmonic u of degree $n \geqq 4$ which is positive in a circle β of radius 33° on σ. Introduce polar coördinates (r, θ, ϕ) so that the center of β corresponds to $r = 1$, $\theta = 0$. Consider

$$v(r, \theta) = \int_0^{2\pi} u(r, \theta, \phi)\, d\phi$$

$v(r, \theta)$ is a spherical harmonic of degree n, of the form $cr^n P_n(\cos\theta)$, $c > 0$, studied by Legendre.

We have

$$P_\nu(1) = 1, \nu = 0, 1, \cdots; \quad P_0(\xi) \equiv 1, \quad P_1(\xi) \equiv \xi,$$

---

[4] R. M. Robinson, *Bulletin of the American Mathematical Society*, 1938, pp. 115-116.

and the recursion formula

$$(\nu+1)P_{\nu+1}(\xi)-(2\nu+1)\xi P_\nu(\xi)+\nu P_{\nu-1}(\xi)=0, \qquad \nu\geqq 1.$$

Denote by $\xi_\nu<1$ the largest value of ξ for which $P_\nu(\xi)=0$. By the recursion formula $P_{\nu+1}(\xi_\nu)<0$ and

$$0=\xi_1<\xi_2<\xi_3\cdots.$$

Furthermore

$$8P_4(\xi)=35\xi^4-30\xi^2+3,$$

$$\xi_4{}^2=\frac{15+\sqrt{120}}{35}>\frac{15+10.6}{36}\geqq\frac{16^2\cdot 10}{6^2\cdot 100}\geqq 0.843^2.$$

Thus we have for $0<\theta_4<90°$ and $\xi_4=\cos\theta_4$ the inequality

$$\cos\theta_4>0.843, \qquad \theta_4<33°.$$

Consequently, for $\xi_\nu=\cos\theta_\nu$, $0<\theta_\nu<90°$,

$$33°>\theta_4>\theta_5>\cdots.$$

But this makes impossible our assumption that u and hence $v(r,\theta)$ are positive in β.

The comparison of Theorem 1 with Lemmas 2 and 3 easily yields Theorem 2.

BERKELEY, CALIFORNIA.

ON THE CHARACTER OF CERTAIN ENTIRE FUNCTIONS IN DISTANT PORTIONS OF THE PLANE.*

By C. V. Newsom.

Recently Ford has proved a theorem pertaining to the behavior of the series,

$$\sum_{n=0}^{\infty} g(n)z^n;\ \text{radius of convergence} = \infty,$$

when the absolute value of z becomes large. In his study the coefficient $g(n)$, when considered as a function $g(w)$ of the complex variable $w = x + iy$, must satisfy the two following conditions throughout any arbitrary right-half plane $x > x_0$:

(a) is single valued and analytic;
(b) is such that for all $|y|$ sufficiently large one may write

$$|g(x+iy)| < Ke^{(\pi+\epsilon)|y|},$$

where ϵ is an arbitrarily small positive quantity given in advance and where K depends only upon x_0 and ϵ.[1]

The present paper studies a similar problem in which, however, condition (b) given above is somewhat less restricted. The theorem to be established is as follows:

THEOREM. *Let it be assumed that the coefficient $g(n)$ occurring in the general term of the power series,*

$$(1)\qquad \sum_{n=0}^{\infty} g(n)z^n;\ \textit{radius of convergence} = \infty,$$

may be regarded as a function $g(w)$ of the complex variable $w = x + iy$ and as such satisfies the following two conditions:

(a) *is single valued and analytic throughout the finite w-plane,*
(b) *is such that for all values of x and y one may write*

---

* Received December 14, 1937.

[1] Ford, *The Asymptotic Developments of Functions Defined by Maclaurin Series,* (Scientific Series, University of Michigan, vol. 11), 1936, pp. 30-37. For a further discussion of developments in this field, see the same treatise.

$$(2) \qquad |g(x+iy)| < Ke^{k\pi|y|},$$

where K is a constant independent of x and y, and k is any given positive integer.

Then the function $f(z)$ defined by series (1) when considered for all values of z satisfying the condition, $-\pi < \arg \pm z < \pi$, may be expressed in the form

$$(3) \quad f(z) = \int_{-l-\frac{1}{2}}^{\infty} \left\{ g(x)[\pm z]^x \frac{\sin k\pi x}{\sin \pi x} \right\} dx - \sum_{m=-l}^{-1} g(m) z^m + \xi_k(l, z);$$

wherein l is any arbitrary positive integer and the upper or lower of the signs $\pm$ is to be taken according as k is odd or even. Moreover, the expression $\xi_k(l, z)$ is such that

$$(4) \qquad \lim_{|z| \to \infty} z^l \xi_k(l, z) = 0,$$

irrespective of the value chosen for l.

Proof of Theorem. The proof of this theorem is based upon the following statement which is seen to be a consequence of Cauchy's integral theorem in the calculus of residues; namely, if $P(w)$ and $Q(w)$ are any two functions of the complex variable $w = x + iy$ both of which are single valued and analytic throughout a region A of the w-plane and of which $Q(w)$ vanishes within A only at the points $w = \lambda_1, \lambda_2, \cdots, \lambda_n$ which are zeros of the first order, and if c_n denotes any closed contour lying within A and including the points $w = \lambda_1, \lambda_2, \cdots, \lambda_n$, then one may write

$$(5) \qquad \frac{1}{2\pi i} \int_{c_n} \frac{P(w)}{[Q(w)]^{k+1}}\, dw = \sum_{n=1}^{n} \frac{1}{k!} \left\{ \frac{d^k}{dw^k} \left[P(w) \left(\frac{w - \lambda_n}{Q(w)} \right)^{k+1} \right] \right\}_{w=\lambda_n},$$

in which k denotes any positive integer. The general term of the summation upon the right is the residue at $w = \lambda_n$ of the integrand upon the left as may be readily verified by observation of its Laurent's expansion.

If the formula of Leibnitz be employed upon the derivative of the product within the bracket of (5), it follows that

$$(6) \quad \frac{1}{2\pi i} \int_{c_n} \frac{P(w)}{[Q(w)]^{k+1}}\, dw$$
$$= \sum_{n=1}^{n} \frac{1}{k!} \sum_{r=0}^{k} \frac{k!}{r!(k-r)!} \left[\frac{d^{k-r}}{dw^{k-r}} P(w) \right]_{w=\lambda_n} \left[\frac{d^r}{dw^r} \left(\frac{w - \lambda_n}{Q(w)} \right)^{k+1} \right]_{w=\lambda_n}.$$

For the proof of the theorem of this paper, we shall choose $Q(w) = \sin \pi w$.

With this choice, $\lambda_n = n$ where n is any positive integer although the analysis follows also when the region A is extended to include a set of negative integers.

Through the medium of elementary transformations, it is found that

$$\left[\frac{d^r}{dw^r}\left(\frac{w-n}{\sin \pi w}\right)^{k+1}\right]_{w=n} = \frac{(-1)^{kn+n}\pi^r}{\pi^{k+1}}\left[\frac{d^r}{dw^r}\left(\frac{w}{\sin w}\right)^{k+1}\right]_{w=0}.$$

Hence it follows at once that

$$(7) \qquad \left[\frac{d^r}{dw^r}\left(\frac{w-n}{\sin \pi w}\right)^{k+1}\right]_{w=n} = \frac{(-1)^{kn+n}\pi^r(k-r)!\,r!}{\pi^{k+1}\cdot k!}\Sigma\alpha_1\alpha_2\cdots\alpha_r,$$

where $\Sigma\alpha_1\alpha_2\cdots\alpha_r$ denotes the sum of the $\binom{k}{r}$ products of r factors each formed by taking the possible combinations of the k quantities, $\pm(k-1)i$; $\pm(k-3)i; \cdots, \left\{\begin{matrix}\pm i\\ 0\end{matrix}\right\}$, r at a time; i having the usual interpretation as the imaginary unit, and where $\left\{\begin{matrix}\pm i\\ 0\end{matrix}\right\}$ is understood as $\pm i$ or 0 according as k is even or odd.[2]

Thus the general term of the summation in (6) with respect to n becomes:

$$\frac{(-1)^{kn+n}}{k!\,\pi^{k+1}}\sum_{r=0}^{k}[(\Sigma\,\alpha_1\alpha_2\cdots\alpha_r)\pi^r P^{(k-r)}(n)],$$

where $P^{(k-r)}(n)$ indicates the $(k-r)$ derivative of $P(w)$ evaluated at $w=n$.

It is now desired that this expression should equal $\frac{g(n)z^n}{k!\,\pi^{k+1}}$ where $g(n)z^n$ is the n-th term of the given series (1). In order that this may be the case it is evidently necessary and sufficient that the hitherto undetermined function $P(n)$ shall satisfy the following differential equation in n:

$$(8) \qquad \sum_{r=0}^{k}[(\Sigma\,\alpha_1\alpha_2\cdots\alpha_r)\pi^r P^{(k-r)}(n)] = g(n)[\pm z]^n,$$

wherein the upper or lower of the signs $\pm$ is to be taken according as k is odd or even.

This non-homogeneous differential equation is well adapted to elementary methods of solution. It has constant coefficients and the roots of its characteristic equation are distinct. Moreover, the roots are observed to be the quantities

$$(9) \qquad \pm(k-1)\pi i, \ \pm(k-3)\pi i, \cdots, \left\{\begin{matrix}\pm \pi i\\ 0\end{matrix}\right\}.$$

---

[2] For a proof of this fact see Newsom, *The American Mathematical Monthly*, vol. 38, no. 9, pp. 500-504.

Henceforth these roots will be denoted by $r_1\pi i$, $r_2\pi i$, etc. where $r_j = k + 1 - 2j$; $1 \leq j \leq k$. Then by a well known formula, a particular solution of (8) may be written directly as follows:

$$(10) \qquad P(n) = \sum_{j=1}^{k} b_j e^{r_j \pi i n} \int_0^n e^{-r_j \pi i n} g(n) [\pm z]^n dn,$$

where

$$(11) \qquad b_j = \frac{1}{(\pi i)^{k-1}} \prod_{a=1}^{j-1} \frac{1}{(r_j - r_a)} \prod_{a=j+1}^{k} \frac{1}{(r_j - r_a)}; \; 1 \leq j \leq k;$$

in which

$$\prod_{a=p}^{q} \frac{1}{(r_j - r_a)} = 1 \text{ if } p \geqq q.$$

Having determined $P(n)$ in this manner, we observe that if n be now replaced by the complex variable $w = x + iy$ the result is an analytic function $P(w)$ of w throughout the finite w-plane.

If then in (6) the path of integration c_n be chosen as any closed (finite) contour which encloses the points $w = -l, -l + 1, -l + 2, \cdots, -1, 0, 1, 2, \cdots, n$; l being any given positive integer; and if we employ the indicated form of the function $P(w)$, equation (6) becomes

$$(12) \qquad \frac{1}{2\pi i} \int_{c_n} \frac{P(w)}{[Q(w)]^{k+1}} dw = \frac{1}{k!\,\pi^{k+1}} \sum_{n=-l}^{n} g(n) z^n.$$

With the actual forms for $P(w)$ and $Q(w)$ substituted into (12), the relation becomes

$$(13) \qquad \sum_{n=0}^{n} g(n) z^n = \frac{k!\,\pi^k}{2i} \sum_{j=1}^{k} b_j \int_{c_n} \frac{e^{r_j \pi i w} \int_0^w e^{-r_j \pi i w} g(w) [\pm z]^w dw}{(\sin \pi w)^{k+1}} dw - \sum_{m=-l}^{-1} g(m) z^m.$$

Thus we arrive at an expression in the form of a contour integral for the sum of the first n terms of the given series (1).

The study of the right-hand member of (13) involves a consideration of k integrals of the form,

$$(14) \qquad \int_{c_n} \frac{e^{r_j \pi i w} \int_0^w e^{-r_j \pi i w} g(w) [\pm z]^w dw}{(\sin \pi w)^{k+1}} dw.$$

Within this contour integral we shall designate

$$(15)\qquad I_j(w, z) = \int_0^w e^{-r_j\pi i w} g(w)\,[\pm z]^w dw.$$

Furthermore, throughout the discussion which follows $[\pm z]$ will be regarded as in the form,

$$[\pm z] = \rho(\cos\phi + i\sin\phi).$$

Moreover, $[\pm z]^w$, when evaluated for any special values of z and $w = x + iy$, is rendered precise in meaning through the following convention:

$$(16)\qquad [\pm z]^w = \rho^x e^{-\phi y} e^{(y\log\rho + x\phi)i}; \qquad -\pi < \phi \leqq \pi.$$

Then it follows that

$$(17)\qquad I_j(w, z) = \int_0^w \rho^x e^{(r_j\pi-\phi)y} e^{[y\log\rho+(\phi-r_j\pi)x]i} g(w)\,dw.$$

For the path of integration to be employed in (17) let us choose that one which, starting at the origin $w = 0$, proceeds along the real axis to the point $w = x$ and then proceeds parallel to the pure imaginary axis to the point $w = x + iy$ in question. Then $I_j(w, z)$ takes the form,

$$(18)\qquad I_j(w, z) = R_j(\rho, \phi, x) + S_j(\rho, \phi, x, y),$$

where

$$(19)\qquad R_j(\rho, \phi, x) = \int_0^x \rho^x e^{(\phi-r_j\pi)xi} g(x)\,dx$$

and

$$(20)\qquad S_j(\rho, \phi, x, y) = i\rho^x \int_0^y e^{(r_j\pi-\phi)y} e^{[y\log\rho+(\phi-r_j\pi)x]i} g(x + iy)\,dy.$$

By virtue of condition (2) postulated upon $g(w)$, if $y > 0$, it is known that

$$|\,g(x + iy)\,| < Ke^{k\pi y}.$$

Thus it follows that

$$(21)\qquad \begin{aligned} |\,S_j(\rho, \phi, x, y)\,| &< K\rho^x \int_0^y e^{[k\pi+r_j\pi-\phi]y} dy \\ &< K\rho^x \frac{e^{[k\pi+r_j\pi-\phi]y}}{[k\pi + r_j\pi - \phi]}, \end{aligned}$$

if the equality sign given in the range of ϕ in (16) be eliminated. In similar manner, if $y < 0$, we have

$$(22)\qquad |\,S_j(\rho, \phi, x, y)\,| < K\rho^x \frac{e^{[-k\pi+r_j\pi-\phi]y}}{|[-k\pi + r_j\pi - \phi]|}.$$

Hence if, in particular, we confine z to any finite region T of the z-plane such that ρ is bounded and within which $-\pi+\epsilon<\phi<\pi-\epsilon$, where ϵ is an arbitrarily small positive quantity, and confine x and y to any strip of finite width (thus making x bounded) drawn in the w-plane parallel to the y-axis, we see that, according as $y>0$ or $y<0$ we may write

$$\begin{aligned} &|S_j(\rho,\phi,x,y)|<Me^{(k\pi+\pi+r_j\pi-\epsilon)y}\\ &|S_j(\rho,\phi,x,y)|<Me^{(-k\pi-\pi+r_j\pi-\epsilon)y}, \end{aligned} \tag{23}$$

where M is an assignable positive constant, independent of ρ, ϕ x, and y. Moreover, for ρ, ϕ, and x as thus bounded, observation of (19) reveals that

$$|R_j(\rho,\phi,x)|<N, \tag{24}$$

where N is a constant independent of ρ, ϕ, and x.

The foregoing properties being premised in regard to the function $I_j(w,z)$, let us return to a further study of the integral (14). For the contour c_n let us take specifically (Compare with the remark preceding (12)) the rectangle formed in the w-plane by the lines,

$$w=-l-\tfrac{1}{2}+iy,\quad w=n+\tfrac{1}{2}+iy,\quad w=x+ip,\ \text{and}\ w=x+iq, \tag{25}$$

where l is an arbitrarily large positive integer, n is an arbitrarily large positive even integer, and p and q may be regarded respectively as any arbitrarily large positive and negative quantities.

We proceed now to consider the four contributions to the integral (14) as the indicated integration over c_n is performed in turn along the four sides of the rectangle described in (25).

First, along the side upon which $w=x+ip$ we have $dw=dx$ and $\sin\pi w=\sin\pi(x+ip)=\sinh\pi p(\sin\pi x\operatorname{ctnh}\pi p+i\cos\pi x)$. Moreover, the integration takes place from $x=n+\frac{1}{2}$ to $x=-l-\frac{1}{2}$. Hence, if the contribution to (14) from this side of c_n be designated by A_j, we have

$$A_j=\int_{n+\frac{1}{2}}^{-l-\frac{1}{2}}\frac{e^{r_j\pi i(x+ip)}[R_j(\rho,\phi,x)+S_j(\rho,\phi,x,p)]}{[\sinh\pi p(\sin\pi x\operatorname{ctnh}\pi p+i\cos\pi x)]^{k+1}}\,dx, \tag{26}$$

where $R_j(\rho,\phi,x)$ and $S_j(\rho,\phi,x,p)$ are defined by (19) and (20). By virtue of relations (23) and (24) it is known that for all the values of ρ, ϕ, and x with which we are concerned in (26) we may write

$$|R_j(\rho,\phi,x)|<N,\quad |S_j(\rho,\phi,x,p)|<Me^{(k\pi+\pi+r_j\pi-\epsilon)p};\ \text{ where }\ \epsilon>0.$$

Hence, upon writing in (26) $r_j\pi i(x+ip)=r_j\pi xi-r_j\pi p$, it appears that the absolute value of the numerator of the integrand of (26) cannot become infinite to a higher order than that of $e^{(k\pi+\pi-\epsilon)p}$ as $p \rightarrow +\infty$. The absolute value of the denominator, however, becomes infinite to as high an order as that of $e^{(k\pi+\pi)p}$ as $p \rightarrow +\infty$ if one recalls the meaning of $\sinh \pi p$ and notes that the quantity in parentheses within the denominator has an absolute value which approaches the limit unity, this being true uniformly for all values of x. Thus we conclude that

$$\lim_{p\rightarrow+\infty} A_j = 0.$$

Similarly, the contribution B_j to the integral (14) arising from the side of c_n upon which $w = x + iq$ $(q < 0)$ is found to approach the limit zero as $q \rightarrow -\infty$. In fact, this contribution is seen to be similar to (26) in form. The integration, however, in this case takes place from $x = -l - \frac{1}{2}$ to $x = n + \frac{1}{2}$ and the expression $S_j(\rho, \phi, x, q)$ which now occurs is related to the second of relations (23) instead of to the first, thus making the absolute value of the numerator of the integrand become infinite to no higher order than that of $e^{(-k\pi-\pi+\epsilon)q}$ as $q \rightarrow -\infty$, while the denominator becomes infinite like $e^{(-k\pi-\pi)q}$ as $q \rightarrow -\infty$.

Next, let us consider the contribution arising from the side $w = n + \frac{1}{2} + iy$. Here $dw = i\,dy$ and, inasmuch as n has been chosen as an even integer, we have $(\sin \pi w)^{k+1} = (\cosh \pi y)^{k+1}$. Having taken $p = +\infty$ and $q = -\infty$, the contribution to (14) from the side in question may be written

$$C_j = ie^{\frac{1}{2}r_j\pi i} \left\{ R_j(\rho, \phi, n + \tfrac{1}{2}) \int_{-\infty}^{\infty} \frac{e^{-r_j\pi y}}{(\cosh \pi y)^{k+1}}\, dy + \int_{-\infty}^{\infty} \frac{e^{-r_j\pi y} S_j(\rho, \phi, n + \frac{1}{2}, y)}{(\cosh \pi y)^{k+1}}\, dy \right\}. \tag{27}$$

The integral,

$$\int_{-\infty}^{\infty} \frac{e^{-r_j\pi y}}{(\cosh \pi y)^{k+1}}\, dy,$$

which we shall denote by H_j may be evaluated directly. It evidently may be rewritten in the form

$$H_j = \frac{2^{k+1}}{\pi} \int_{0}^{\infty} \frac{t^{k-r_j}}{(t^2 + 1)^{k+1}}\, dt,$$

as appears by making the transformation $e^{\pi y} = t$. This new integral may be evaluated at once and is equal to

$$(28)\quad H_j = \int_{-\infty}^{\infty} \frac{e^{-r_j \pi y}}{(\cosh \pi y)^{k+1}}\, dy = \frac{(2i)^{k+1} i^{-r_j}}{(r_j - k - 1)\pi} \cdot \prod_{\beta=1}^{j-1}\left(1 - \frac{k - r_j + 1}{2\beta}\right) \prod_{\beta=j+1}^{k}\left(1 - \frac{k - r_j + 1}{2\beta}\right), \quad 1 \leq j \leq k;$$

in which

$$\prod_{\beta=p}^{q}\left(1 - \frac{k - r_j + 1}{2\beta}\right) = 1 \text{ if } p \geqq q.^{3}$$

As to the second integral in (27) we observe from (21) and (22) and the paragraph which follows them that the absolute value of the integrand is less than

$$(29)\quad \frac{M_1 \rho^{n+\frac{1}{2}} e^{[k\pi+\pi-\epsilon]y}}{(\cosh \pi y)^{k+1}} \quad \text{or} \quad \frac{M_1 \rho^{n+\frac{1}{2}} e^{[-k\pi-\pi+\epsilon]y}}{(\cosh \pi y)^{k+1}},$$

where M_1 is an assignable positive constant independent of ρ, ϕ, n, and y; the first or the second of these expressions to be used according as $y > 0$ or $y < 0$. In view of the fact that $(\cosh \pi y)^{k+1}$ becomes infinite like $e^{(k\pi+\pi)y}$ as $y \to +\infty$ and like $e^{(-k\pi-\pi)y}$ as $y \to -\infty$, it follows that this integral converges and converges uniformly for all values of z lying in the region T of the z-plane as described immediately under expression (22). Furthermore, we observe from (29) that if ρ be restricted to values less than unity, that is, if we consider only that portion of the region T which lies within the unit circle about the origin, then the limit of the integral in question as $n \to \infty$ will be 0. Thus, in conclusion, we may write

$$(30)\quad C_j = i^{(1+r_j)} H_j R_j(\rho, \phi, n + \tfrac{1}{2}) + \eta_j(n, z),$$

where for the values of z under consideration, namely, for those in T such that $|z| < 1$, it follows that

$$(31)\quad \lim_{n\to\infty} \eta_j(n, z) = 0.$$

There now remains to be considered one contribution to the integral (14), namely, the one arising from the integration over the side $w = -l - \frac{1}{2} + iy$ of the rectangle c_n. This, however, is merely what is obtained just above by using $-l - \frac{1}{2}$ instead of $n + \frac{1}{2}$ and noting that the integration now takes place from $+\infty$ to $-\infty$ instead of vice-versa. Thus this contribution takes the form,

$$(32)\quad D_j = -i^{(1+r_j)} H_j R_j(\rho, \phi, -l - \tfrac{1}{2}) + Y_j(l, z),$$

---

[3] For this evaluation note formula 20, table 17, in the integral table of Bierens de Haan. It should be observed, however, that this formula introduces a removable indeterminate form when r_j is replaced by its proper value.

where

$$(33) \qquad Y_j(l, z) = -i^{(1+r_j)} \int_{-\infty}^{\infty} \frac{e^{-r_j\pi i} S_j(\rho, \phi, -l-\tfrac{1}{2}, y)}{(\cosh \pi y)^{k+1}}\, dy.$$

As regards the integral $Y_j(l, z)$ here encountered it should be observed that an analysis similar to that which was applied to the second integral of (27) may likewise be employed here. In fact, the integrand of this integral is less in absolute value than

$$\frac{M_2 \rho^{-l-\frac{1}{2}} e^{[k\pi+\pi-\epsilon]y}}{(\cosh \pi y)^{k+1}} \quad \text{or} \quad \frac{M_2 \rho^{-l-\frac{1}{2}} e^{[-k\pi-\pi+\epsilon]y}}{(\cosh \pi y)^{k+1}},$$

according as $y > 0$ or $y < 0$. Thus it follows that $Y_j(l, z)$ converges uniformly throughout the portion of T in which $|z| = \rho \geqq \rho_1 > 0$, where ρ_1 is arbitrarily small; and, furthermore, is such that for those values of z in the portion of T just specified we shall have

$$(34) \qquad \lim_{|z| \to \infty} z^l Y_j(l, z) = 0.$$

We now recall from (19) that

$$R_j(\rho, \phi, x) = \int_0^x \rho^x e^{(\phi - r_j\pi)xi} g(x)\, dx,$$

in which $\rho e^{i\phi} = \pm z$, the sign $+$ or $-$ being used according as k is odd or even. Hence we may write

$$R_j(\rho, \phi, n+\tfrac{1}{2}) - R_j(\rho, \phi, -l-\tfrac{1}{2}) = \int_{-l-\frac{1}{2}}^{n+\frac{1}{2}} g(x)[\pm z]^x e^{-r_j\pi xi}\, dx.$$

If we now add the various contributions A_j, B_j, C_j, and D_j which make up the integral (14), we conclude that

$$(35) \qquad \int_{C_n} \frac{e^{r_j\pi i w} \int_0^w e^{-r_j\pi i w} g(w)[\pm z]^w dw}{(\sin \pi w)^{k+1}}\, dw$$
$$= i^{(1+r_j)} H_j \int_{-l-\frac{1}{2}}^{n+\frac{1}{2}} g(x)[\pm z]^x e^{-r_j\pi xi}\, dx + \eta_j(n, z) + Y_j(l, z),$$

wherein $\eta_j(n, z)$ and $Y_j(l, z)$ possess the properties indicated respectively in (31) and (34).

It is now possible, therefore, to restate (13) in the form,

$$(36)\qquad \sum_{n=0}^{n} g(n) z^n = \sum_{j=1}^{k} \frac{k!\,\pi^k}{2i} \cdot i^{(1+r_j)} b_j H_j$$
$$\int_{-l-\frac{1}{2}}^{n+\frac{1}{2}} g(x)[\pm z]^x e^{-r_j \pi x i} dx + \sum_{j-1}^{k} \frac{k!\,\pi^k}{2i} b_j \eta_j(n, z)$$
$$+ \sum_{j=1}^{k} \frac{k!\,\pi^k}{2i} b_j Y_j(l, z) - \sum_{m=-l}^{-1} g(m) z^m.$$

If we recall the value of b_j given in (11) and the value of H_j given in (28), it is a matter of simple algebra to determine that the entire coefficient preceding the integral within the first summation upon the right is equal to unity. Moreover it is readily shown that

$$\sum_{j=1}^{k} e^{-r_j \pi x i} = \frac{\sin k\pi x}{\sin \pi x}.$$

Hence it follows at once that

$$(37)\qquad \sum_{n=0}^{n} g(n) z^n = \int_{-l-\frac{1}{2}}^{n+\frac{1}{2}} \left\{ g(x)[\pm z]^x \frac{\sin k\pi x}{\sin \pi x} \right\} dx$$
$$- \sum_{m=-l}^{-1} g(m) z^m + \eta(n, z) + \xi_k(l, z),$$

where $\eta(n, z)$ and $\xi_k(l, z)$ possess the properties previously held respectively by $\eta_j(n, z)$ and $Y_j(l, z)$; that is, for all z in T having $|z| < 1$ we may write

$$\lim_{n \to \infty} \eta(n, z) = 0,$$

and for all z in T for which $|z| \geqq \rho_1 > 0$, $\xi_k(l, z)$ converges uniformly and

$$\lim_{|z| \to \infty} z^l \xi_k(l, z) = 0.$$

If we now confine z to values in T for which $0 < \rho_1 \leqq |z| < 1$ and then allow $n \to \infty$ in the relation (37), it follows from what has been said of $\eta(n, z)$ that we may write for all such z:

$$(38)\qquad f(z) = \sum_{n=0}^{\infty} g(n) z^n = \int_{-l-\frac{1}{2}}^{\infty} \left\{ g(x)[\pm z]^x \frac{\sin k\pi x}{\sin \pi x} \right\} dx$$
$$- \sum_{m=-l}^{-1} g(m) z^m + \xi_k(l, z).$$

In this equation, however, the left member is an analytic function of z for all finite z, as follows from the original hypothesis that the radius of convergence of the series is infinite. Likewise, the right member is an analytic function of z for all z in T for which $|z| \geqq \rho_1 > 0$ as follows especially from the observation previously made that $\xi_k(l, z)$ converges uniformly for all such z. It follows, therefore, as the result of well known principles of analytic continuation that (38) holds true for all z in T for which $|z| \geqq \rho_1 > 0$, and the theorem as originally stated has been established.

Remarks and generalizations. It is desirable to note that in order to secure simplicity of statement, the condition (a) pertaining to the coefficient $g(n)$ of the theorem was made unnecessarily restrictive. In fact, if one examines the proof it appears readily that the following generalization may be made; namely, in case condition (a) is not satisfied but instead the function $g(w)$, while still remaining single valued throughout the finite w-plane, has m $(m \geqq 1)$ singularities situated at the points $w = w_1, w_2, \cdots, w_m$, none of which are negative integers, then (3) continues to hold true provided one subtracts from the right member the sum of the residues of the function $P(w)/[Q(w)]^{k+1}$, at these points.

The same conclusion applies also in case one or more of the points w_p are negative integers, as $w = -q$ $(q \geqq 1)$, except that the alterations taking place on the right in (3) must then include the deletion of the term $g(-q)/z^q$ occurring in the summation.

Also it should be remarked that condition (b) pertaining to the coefficient $g(n)$ of the theorem may be replaced by an alternate condition, namely, that $g(w)$ can be such that for all values of x and y one may write

$$\left| \frac{g(x + \frac{1}{2} + iy)}{g(x)} \right| < Ke^{k\pi|y|},$$

where K is a constant independent of x and y and k is any given positive integer $\geqq 1$. With this new condition (b) and with the other conditions as originally specified, conclusion (3) follows as before. Usually the original condition (b) is more readily applied but in certain cases when the coefficient $g(n)$ contains exponentials the alternate condition becomes desirable.

In establishing these latter remarks it should be observed that the indicated alteration upon condition (b) changes the proof of the theorem only slightly. Virtually the only essential change takes place in the expressions of (29) where the absolute value of the integrand in question now is less than

$$\frac{M_1 \,|\, g(n) \,|\, \rho^n e^{[k\pi+\pi-\epsilon]y}}{(\cosh \pi y)^{k+1}} \quad \text{or} \quad \frac{M_1 \,|\, g(n) \,|\, \rho^n e^{[-k\pi-\pi+\epsilon]y}}{(\cosh \pi y)^{k+1}},$$

where M_1 is an assignable positive constant independent of ρ, ϕ, n, and y; the first or the second of these expressions to be used according as $y > 0$ or

$y < 0$. As before, the second integral in (27) converges uniformly for the values of z lying in such a region of the z-plane as described immediately after (22). But $|g(n)|\rho^n$ is the absolute value of the n-th term of the original convergent power series (1) and thus approaches zero as $n \rightarrow +\infty$. So, without any additional restrictions upon ρ, the integral in question approaches zero as $n \rightarrow +\infty$. The remainder of the proof of the theorem readily follows without the necessity of employing the device of analytic continuation.

Remarks concerning the applications of the theorem. Due to the presence of l in the lower limit of the integral of (3), the theorem of this paper does not furnish direct information upon the asymptotic representation of the given function, $f(z)$. However, it is in regard to such asymptotic studies that the theorem appears to have important applications. These applications will apparently follow the same general procedures as previously discussed by Ford and Van Engen.[4] Since, however, for all $x > x_0$, where x_0 is arbitrary, and for all $|y|$ sufficiently large one may write

$$\frac{1}{|\Gamma(x+iy)|} < Ke^{(\pi/2+\epsilon)|y|},$$

where ϵ is an arbitrarily small positive quantity given in advance, and where K depends only upon ϵ and x_0,[5] it follows that condition (b) of the theorem will be satisfied when $g(n)$ is the reciprocal of the product of any number of gamma functions of the type, $\Gamma(n+p)$. Thus in such a case the given function, $f(z)$, may be directly represented as in relation (3).

THE UNIVERSITY OF NEW MEXICO.
ALBUQUERQUE, NEW MEXICO.

---

[4] Ford, *op. cit.*, pp. 63-85.

[5] Ford, *op. cit.*, p. 61.

ON THE REPRESENTATION OF BOUNDED ANALYTIC FUNCTIONS BY SEQUENCES OF POLYNOMIALS.*

By O. J. FARRELL.

1. Introduction. Let the function $f(z)$ be analytic and bounded in a region G of the plane of the complex variable z. Let $d(f, z_0)$ denote the diameter[1] of the cluster set[2] of $f(z)$ at the point z_0 on the boundary B of G. Let $D(f, B)$ denote the maximum of $d(f, z_0)$ for all points z_0 on B. A natural question to be raised is: When can $f(z)$ be represented in G by a sequence of polynomials $p_n(z), n = 1, 2, \cdots$ such that

$$\overline{\lim_{n\to\infty}} \{\overline{\text{bound}} \, [\, | f(z) - p_n(z) |, z \text{ in } G]\} \leqq D(f, B) \,?$$

The present paper contributes toward the solution of this problem the following two theorems.

THEOREM 1. *Let the function $f(z)$ be analytic and bounded in the interior T of a limited Jordan curve J. Then there exists a sequence of polynomials $p_1(z), p_2(z), \cdots$ converging to $f(z)$ in T, uniformly on any closed set in T, so that*

$$\overline{\lim_{n\to\infty}} \{\overline{\text{bound}} \, [\, | f(z) - p_n(z) |, z \text{ in } T]\} \leqq D(f, J). \tag{1}$$

THEOREM 2. *In order that corresponding to* EVERY *function $f(z)$ analytic and bounded in a region G there shall exist a sequence of polynomials $p_1(z), p_2(z), \cdots$ converging to $f(z)$ in G so that*

$$\overline{\lim_{n\to\infty}} \{\overline{\text{bound}} \, [\, | f(z) - p_n(z) |, z \text{ in } G]\} \leqq D(f, B) \tag{2}$$

it is necessary that G be finite and that $G + B$ be the complement of an infinite region. Moreover, if G is simply connected it is necessary that every point of the boundary of G be also a boundary point of the infinite region.

---

* Presented to the American Mathematical Society, September 7, 1937. Received by the Editors October 25, 1937; Revised January 6, 1938.

[1] The usual definition of diameter is meant here, namely the maximum distance between any two points of the cluster set.

[2] See Seidel, "On the cluster values of analytic functions," *Transactions of the American Mathematical Society*, vol. 34 (1932), pp. 1-21.

2. Proof of Theorem 1. Let T be enclosed in a sequence of limited simply connected regions $T_1, T_2, \cdots$ all containing $T+J$ and such that every region T_{n+1} together with its boundary lies interior to T_n while no point exterior to T lies in all the T_n. Let $z=\alpha$ be a fixed point in T and let the function $\psi_n(z)$ map the region T_n conformally onto the region T so that the point $z=\alpha$ corresponds to itself and so that $\psi'_n(\alpha)>0$. Then we have[3] uniformly in $T+J$

$$\lim_{n\to\infty} \psi_n(z) = z.$$

Let $p_n(z)$ denote a polynomial which approximates the function

$$f_n(z) = f[\psi_n(z)]$$

so that[4]

$$| p_n(z) - f_n(z) | \leqq 1/n, \qquad z \text{ in } T+J.$$

Thus we have for z in T, uniformly on any closed set in T,

$$\lim_{n\to\infty} p_n(z) = f(z).$$

We shall prove that these polynomials also fulfill the further requirements of the theorem. We find it convenient to do this in two steps as follows. We prove first that if z_0 be any point of J, if $z_1, z_2, \cdots$ be any sequence of points in G approaching z_0 as limit, and if $p_{\mu_1}(z), p_{\mu_2}(z), \cdots$ be any subsequence of the polynomials $p_n(z)$, then

$$\overline{\lim_{n\to\infty}} \, | f(z_n) - p_{\mu_n}(z_n) | \leqq d(f, z_0). \tag{3}$$

The second part of the proof is by way of contradiction, for we show that if the polynomials $p_n(z)$ do not fulfill (1) they violate (3).

We proceed with the proof of (3). Let z_0 be any point of J and let S_{z_0} denote the cluster set of $f(z)$ at z_0. If a positive ϵ be assigned, then for all n greater than a certain N we have simultaneously

(a) z_n and $\psi_{\mu_n}(z_n)$ both so near to z_0 that the image points $f(z_n)$ and $f_{\mu_n}(z_n)$ are both within distance ϵ of S_{z_0};

(b) $| p_{\mu_n}(z_n) - f_{\mu_n}(z_n) | \leqq \epsilon.$

---

[3] See Walsh, "Interpolation and approximation by rational functions in the complex domain," *Colloquium Publications of the American Mathematical Society*, vol. 20 (1935), p. 35, Corollary 1.

[4] The function $f_n(z)$ is analytic in T_n and hence by Runge's theorem can be represented in T_n by a series of polynomials converging uniformly on any closed point set (such as $T+J$) lying wholly interior to T_n.

Consequently we have

$$|f(z_n) - p_{\mu_n}(z_n)| \leqq d(f, z_0) + 3\epsilon, \qquad n > N,$$

which proves (3).

Suppose now that (1) does not hold. Then there must exist a subsequence $p_{\lambda_1}(z), p_{\lambda_2}(z), \cdots$ of the $p_n(z)$ such that for EVERY $p_{\lambda_n}(z)$ we have

$$\overline{\text{bound}}\,[\,|f(z) - p_{\lambda_n}(z)|, z \text{ in } T] \geqq L > D(f, J).$$

Associate with each polynomial $p_{\lambda_n}(z)$ a point z_n in T within distance $1/n$ of J and at the same time so near to J that

$$|f(z_n) - p_{\lambda_n}(z_n)| > \overline{\text{bound}}\,[\,|f(z) - p_{\lambda_n}(z)|, z \text{ in } T] - 1/n.$$

Let $z'_1, z'_2, \cdots$ be a subsequence of the points z_n approaching a point z_0 on J as limit. Let $p_{\mu_n}(z)$ denote that polynomial of the sequence $p_{\lambda_n}(z)$ with which the point z'_n has been associated. Then for the sequence z'_n we have

$$\overline{\lim_{n\to\infty}}\,|f(z'_n) - p_{\mu_n}(z'_n)| \geqq L > D(f, J) \geqq d(f, z_0).$$

This contradicts (3) and completes the proof of the theorem.

3. Proof of Theorem 2. It follows at once from the boundedness of $f(z)$ in G that polynomial representation of $f(z)$ in G characterized by (2) requires that G be finite.

We note next that among all the functions $f(z)$ that are bounded in G are included all the functions that are analytic in $G + B$. Moreover, if $f(z)$ is analytic in $G + B$, then $D(f, B) = 0$; and any sequence of polynomials $p_n(z)$ which fulfills (1) thereby converges to $f(z)$ uniformly in $G + B$. But a necessary condition that every function analytic on a closed limited point set can be uniformly approximated thereon by a polynomial is that the point set be the complement of an infinite region.[5] In the present situation, where the point set in question consists of a region and its boundary, the condition that the point set be the complement of an infinite region is equivalent to the condition that B divide the extended plane into precisely two regions: G and an infinite region H.

When G is simply connected a further condition is necessary for the representation of a function $f(z)$ analytic and bounded in G by polynomials such that (2) holds. This condition is that B shall be the common boundary of G and H, or (what is the same) that the neighborhood of every point of B

---

[5] Walsh, *op. cit.*, p. 26, Theorem 17.

shall contain points of H. To prove this suppose there are points of B in the neighborhood of which there are no points of H. Choose for $f(z)$ any function which maps G conformally onto the interior of the unit circle. Let C denote the portion of B that bounds H and let R denote the finite region bounded by C. If polynomials $p_n(z)$ satisfy (2) while converging to $f(z)$ in G it follows that

$$\varlimsup_{n\to\infty} \{\text{Max}\,[\,|\,p_n(z)\,|, z \text{ in } G + B]\} \leqq 1 + D(f, B) \leqq 3.$$

The $p_n(z)$ are therefore uniformly bounded on B and hence also on C and consequently in R as well. Thus they form a normal family in R from which we can extract a subsequence $p_{\lambda_n}(z)$ converging in R to a function $F(z)$ analytic in R and equal to $f(z)$ in G. If ζ be any point of B that lies in R, it follows that $|\,F(\zeta)\,| = 1$; also that $|\,F(z)\,| \leqq 1$ in the neighborhood of ζ. Then by the Principle of Maximum[6] $f(z)$ must be identically constant, which is impossible. Thus the mapping function $f(z)$ can not be represented in G by polynomials for which (2) holds. This completes the proof of Theorem 2.

UNION COLLEGE.

---

[6] See Walsh, *op. cit.*, p. 4.

INVARIANT MANIFOLDS NEAR AN INVARIANT POINT OF UNSTABLE TYPE.*

Daniel C. Lewis, Jr.[1]

1. Introduction and formulation of the problem. In the case of real surface transformations, not necessarily analytic, the existence of an invariant curve in the neighborhood of an unstable invariant point (and containing this invariant point) was proved by Hadamard. In the case of the analytic surface transformations arising in dynamics, similar results were obtained by Poincaré and Birkhoff. Lattés[2] improved the results of Poincaré and also made some use of the methods of Hadamard. He obtained the generalization of Poincaré's result to transformations in three dimensional space. More recently, in the case of the surface transformations of dynamics, Birkhoff[3] has proved the existence of invariant curves of certain kinds near the given invariant point, but which do not contain this invariant point. In this paper we give a generalization of Hadamard's method to the case of transformations in spaces of an arbitrary number of dimensions. We shall, however, be interested primarily in analytic transformations, not because the method is inapplicable to the non-analytic case, but because the hypothesis of analyticity enables us to draw more complete conclusions. The main advantage in adopting Hadamard's method is that it gives a clearer picture of the way in which the invariant manifolds are the uniform limits of the successive images of other manifolds. We shall show that the method holds for complex spaces, and, since any transformation in a real space defined by power series can be interpreted as a trans-

---

* Received January 17, 1938.

[1] These results were obtained for the most part while the writer was a National Research Fellow.

[2] H. Poincaré, *Oeuvres dè, publiées sous les auspices de l'académie des sciences*, vol. 1 (1928), Gauthier Villars, Paris, *Mémoire sur les courbes définies par les équations différentielles*, chapter XIX, especially pp. 202-203; J. Hadamard, "Sur l'itération et les solutions asymptotiques des équations différentielles," *Bulletin de la Société Mathématique de France*, vol. 29 (1901), pp. 224-228; G. D. Birkhoff, "Surface transformations and their dynamical applications," *Acta Mathematica*, vol. 43 (1922), pp. 1-119, especially pp. 45-49; S. Lattés, "Sur les équations fonctionelles qui definissent une courbe ou une surface invariante par une transformation," *Annali di Matematica*, ser. 3, vol. 13 (1907), pp. 1-137.

[3] G. D. Birkhoff, "Nouvelles recherches sur les systèmes dynamiques," *Memoriae, Pontificiae Academiae Scientiae Novi Lyncaei*, series 3, vol. 1, pp. 85-216, especially chapter III.

formation in the corresponding complex space (of twice the number of real dimensions), the analyticity of the invariant manifolds follows at once from well known theorems in the theory of analytic functions of several complex variables. This is because of the fact that these invariant manifolds may be taken as the uniform limits of sequences of analytic manifolds. Details will be given in § 5.

The proofs of Poincaré and Birkhoff use the theory of majorantes and thus also yield analyticity. Birkhoff's proof makes essential use of the hypothesis of "conservativity," a special property enjoyed by transformations arising in dynamics.

We consider a transformation T which with a suitable choice of coördinates can be written in the following form:

$$(1.1)\qquad \begin{aligned} \bar{x}_i &= s_i x_i + f_i(x, y), && (i = 1, 2, \cdots, m) \\ \bar{y}_h &= \sigma_h y_h + \phi_h(x, y), && (h = 1, 2, \cdots, n). \end{aligned}$$

It is supposed that f_i and ϕ_h are convergent power series in $x_1, \cdots, x_m, y_1, \cdots, y_n$ lacking constant and linear terms. As for the complex numbers s_i and σ_h, it is assumed that $| s_i | > | \sigma_h | > 0$. Denoting by s the minimum of the $| s_i |$ and by σ the maximum of the $| \sigma_h |$, we have $| s_i | \geqq s > \sigma \geqq | \sigma_h |$.

Under this transformation a sufficiently small neighborhood of the origin in the $m + n$ dimensional complex space is carried over into another neighborhood in a one-to-one manner, the origin itself being an invariant point. We agree once and for all to restrict attention to a neighborhood common to these two.

It is hardly necessary to note in passing that we are here dealing in a certain sense with the most general type of analytic transformation having a non-vanishing jacobian at an invariant point. In fact the matrix of the jacobian evaluated at the invariant point need only to have simple elementary divisors in order for the transformation to be representable in the form (1.1) in the neighborhood of the invariant point. The only other assumption is that the absolute values of the characteristic roots are not all equal to each other.

Let C be an arbitrary (local) complex "manifold" with the equations $y_h = y_h(x)$ $(h = 1, \cdots, n)$, where the $y_h(x)$ are single valued analytic functions of the x_i defined when each $| x_i |$ is sufficiently small and vanishing when $x_1 = \cdots = x_m = 0$. It is our principal aim to investigate the existence of a manifold C^* of this type every point of which sufficiently near the origin is transformed by T into a point of C^*. Such a manifold will be called invariant.

We shall find that such manifolds do exist if $s > 1$. If $s \leqq 1$, a consideration of the inverse transformation shows the existence of invariant manifolds of the type $x_i = x_i(y)$. Thus, invariant manifolds of some type always exist under our assumption, $s > \sigma$.

2. Preliminary considerations on implicit functions. We begin with the theory of implicit functions, proving a theorem which states that the implicit functions are defined not only in a neighborhood of a point but in a region which for our purposes is sufficiently extensive.

THEOREM 1. *Let $g_i(u, v)$ $(i = 1, \cdots, p)$ be p analytic functions of the $m + p$ complex variables, $u_1, \cdots, u_m, v_1, \cdots, v_p$ defined in a region R given by the inequalities,*

$$|u_h| < A_h, \quad |v_k| < B_k; \qquad (h = 1, \cdots, m;\ k = 1, \cdots, p). \tag{2.1}$$

Let the g_i furthermore satisfy the conditions

$$g_i(0, 0) = 0, \tag{2.2}$$

$$\frac{\partial(g_1, \cdots, g_p)}{\partial(v_1, \cdots, v_p)} \neq 0 \text{ in } R; \tag{2.3}$$

so that the functions F_{kh} defined by the equations

$$\sum_{k=1}^{p} \frac{\partial g_i}{\partial v_k} F_{kh} - \frac{\partial g_i}{\partial u_h} = 0 \tag{2.4}$$

must exist and be analytic in R. Moreover suppose that

$$|F_{kh}| < M_{kh} \text{ in } R \tag{2.5}$$

and let G_h $(h = 1, \cdots, m)$ be any m positive numbers such that

$$\sum_{h=1}^{m} M_{kh} G_h \leqq B_k; \qquad G_h \leqq A_h. \tag{2.6}$$

Then there exists a unique set of analytic functions $v_i(u)$ $(i = 1, \cdots, p)$, of the m complex variables $u_1, \cdots, u_m$, defined in a region S given by

$$|u_h| < G_h, \qquad (h = 1, \cdots, m), \tag{2.7}$$

such that each of the following three conditions is satisfied:

$$|v_k(u)| < B_k \text{ for } (u) \text{ in } S, \qquad (k = 1, \cdots, p). \tag{2.8}$$

$$v_k(0) = 0. \tag{2.9}$$

$$g_k(u, v(u)) \equiv 0 \text{ in } S.^{4} \tag{2.10}$$

Proof. It is readily verified by differentiation of (2.4) and with the help of (2.3), that the F's satisfy the identities

$$\sum_{a=1}^{p} \frac{\partial F_{kh}}{\partial v_a} F_{al} + \frac{\partial F_{kh}}{\partial u_l} \equiv \sum_{a=1}^{p} \frac{\partial F_{kl}}{\partial v_a} F_{ah} + \frac{\partial F_{kl}}{\partial u_h}, \qquad (h, l = 1, \cdots, m). \tag{2.11}$$

We now consider the system of differential equations

$$\frac{\partial v_k}{\partial u_h} = F_{kh}, \tag{2.12}$$

the complete integrability conditions of which are satisfied in virtue of (2.11). Any set of analytic functions satisfying (2.10) must be a solution of (2.12), as we see by differentiating (2.10) and using (2.4). Conversely, it is readily verified that the unique solution of (2.12) which vanishes at the origin will also satisfy (2.10). We shall next show that this solution, which is known to exist in the neighborhood of the origin, may also be defined throughout S.

Let $(\bar{u})$ be an arbitrary point of S and let $u_h = u_h{}^0(t)$ be a straight line L joining the origin with $(\bar{u})$, considered as a point in the Euclidean space of $2m$ real dimensions. The real parameter t may be chosen as the distance of the point $(u^0(t))$ from the origin, so that we have $u_h{}^0(0) = 0$, $u_h{}^0(T) = \bar{u}_h$, $T \geqq 0$. Let the values of the functions $v_k{}^0(t)$ be determined from the existence theorems for ordinary differential equations by the following system:

$$\frac{dv_k{}^0}{dt} = \sum_{h=1}^{m} F_{kh}(u^0(t), v^0) \frac{du_h{}^0}{dt} \qquad (k = 1, \cdots, p) \tag{2.13}$$

together with the initial conditions $v_k{}^0(0) = 0$. The v^0's can be determined as functions of t so long as none of the $|v_k{}^0(t)|$ reaches the value B_k, beyond which the F's are not defined. Cf. (2.1) and (2.4). Now $|v_k{}^0(t)|$ can not increase any faster than it would if the right member of (2.13) were replaced by $\sum_{h=1}^{m} M_{kh} \left| \frac{du_h{}^0}{dt} \right|$ from (2.5). Integrating, we find, on using (2.7) and

---

[4] It is not true, however, that the functions $v_i(u)$ necessarily yield all the solutions of the system of equations $g_i = 0$ in the region $|u_h| < G_h$, $|v_k| < B_k$, as we see from the following elementary example: Let $p = m = 1$ and let

$$g_1(u, v) = -1 + \exp 4\pi i(u_1 - v_1),$$

a function which is certainly both analytic and provided with a non-vanishing derivative with respect to v_1 in any finite region R we may choose to name, e. g. $|u_1| < 2 = A_1$, $|v_1| < 2 = B_1$. It turns out that $F_{11} \equiv 1$ and so we may choose $M_{11} = 2$ and $G = 1$. But in addition to the solutions provided by the theorem, viz. $v_1 = u_1$, the equation $g_1 = 0$ also admits (with infinitely many others) the solution, $u_1 = 0$, $v_1 = 1/2 < 2$.

(2.6) that we would have $|v_k^0(t)| \leqq \sum_{h=1}^{m} M_{kh} |u_h(t)| < B_k$ for $0 \leqq t \leqq T$. Here we have used the fact that L is a straight line and that $(u_h(t))$ is in S for $0 \leqq t \leqq T$. This shows that $v_k^0(t)$ is defined for $0 \leqq t \leqq T$. Now we set $v_k(\bar{u}) = v_k^0(T)$ and thus the v's are defined over S. We have shown at the same time that the v's as thus defined satisfy (2.8) and (2.9). That they also satisfy (2.12), and hence (2.10), and the fact that they are unique, now follows by an obvious adaptation from the classical theory of completely integrable total differential equations as presented by Morera or Mayer.[5]

It may be mentioned in passing that the above theorem is true also for real functions $g_i(u, v)$ of class C'' in the real variables (u) and (v), the solving functions $v_i(u)$ then being of class C'. The same proof is valid, if we understand L now as a straight line in a Euclidean space of m real dimensions.

3. On the transformation of a manifold belonging to $\mathbf{K(\alpha, \beta)}$. We now return to the proper subject matter of the present investigation.

For greater brevity we use x and y without indicial subscripts to denote respectively the sets $(x_1, \cdots, x_m)$ and $(y_1, \cdots, y_n)$; and an x or y capitalized or tagged in some other manner will have a similar interpretation. For example $\bar{y}'$ would stand for the set $(\bar{y}'_1, \cdots, \bar{y}'_n)$. We admit, however, the following exceptions: $|x|$ will denote the largest of the numbers $|x_1|, \cdots, |x_m|$ and $|y|$ will denote the largest of the numbers $|y_1|, \cdots, |y_n|$. A further exception is introduced in § 4.

We use the notation d/dx_i to denote partial differentiation when the y's are functions of the x's. On the other hand $\partial/\partial x_i$ means that all the $m + n$ variables x and y are considered as independent. The symbol $|dy/dx|$ is used to represent the largest of the numbers $|dy_i/dx_h|$; $(i = 1, \cdots, n;\ h = 1, \cdots, m)$.

In considering any point on a manifold of the type of C, described in § 1, the x's will sometimes be called the abscissas and the y's the ordinates of the point in question.

Let us associate with any two positive numbers α and β the class $K(\alpha, \beta)$ of manifolds C such that $y(x)$ is defined and $|dy/dx| < \alpha$ for $|x| < \beta$.

Evidently $K(\alpha, \beta) \subset K(\alpha, \beta')$, if $\beta > \beta'$. We now prove

THEOREM 2. *β can always be chosen so small that any manifold C in $K(\alpha, \beta)$ is transformed by T into a manifold $\bar{C}$ in $K(\alpha, \beta s^*)$, where s^* is a preassigned positive number $< s$.*

---

[5] Cf. Levi-Civita, *The Absolute Differential Calculus*, English translation by M. Long (1927), pp. 22-25.

Proof. The equations of $\bar{C}$ can be written with the parameters x in the form,

$$\begin{aligned} \bar{x}_i &= s_i x_i + f_i(x, y(x)) \\ \bar{y}_h &= \sigma_h y_h(x) + \phi_h(x, y(x)). \end{aligned} \tag{3.1}$$

On elimination of x, these equations take the form $\bar{y}_h = \bar{y}_h(\bar{x})$, and it is our first purpose to show that $\bar{y}(\bar{x})$ is defined for $|\bar{x}| < \beta s^*$ if β is sufficiently small and C belongs to $K(\alpha, \beta)$. This we do by appealing to Theorem 1, taking $p = m$, $\bar{x}_i = u_i$, $x_i = v_i$, $g_i(u, v) \equiv s_i x_i - \bar{x}_i + f_i(x, y(x))$, $B_i = \beta$, $A_h = \infty$. For, since f_i does not contain linear terms in x and y and since $|dy/dx| < \alpha$, we see that all the derivatives of the $f_i(x, y(x))$ with respect to $x_1, \cdots, x_m$ approach zero with $|x|$ uniformly with respect to the manifolds C considered. It follows that for $C \subset K(\alpha, \beta)$ we may take the M_{kh} of Theorem 1 equal to an expression of the form $\delta_{kh} |s_h^{-1}| + \epsilon$, where $\delta_{kh} = 0$ if $k \neq h$ and $\delta_{hh} = 1$, and where ϵ tends uniformly to 0 with β. Hence by taking β sufficiently small, we see from (2.6) that we may take $G_h = \beta s^*$. All this shows that we may uniquely and analytically solve the first m of equations (3.1) for the x's in terms of the $\bar{x}$'s by means of equations of the form $x_i = x_i(\bar{x})$, where the $x_i(\bar{x})$ are defined for $|\bar{x}| < \beta s^*$, and in such manner that $x_i(0) = 0$ and $|x(\bar{x})| < \beta$. In virtue of these last inequalities we can substitute the values of x_i into the last n of equations (3.1) and thus obtain the required equations of $\bar{C}$ in the form $\bar{y}_h = \bar{y}_h(\bar{x})$ defined for $|x| < \beta s^*$, provided again that β is sufficiently small. And the result holds for any C in $K(\alpha, \beta)$.

We next prove that for the functions thus obtained $|d\bar{y}/d\bar{x}| < \alpha$ for $|\bar{x}| < \beta s^*$ and $C \subset K(\alpha, \beta)$, at least for β sufficiently small, i. e. perhaps still smaller than the value already assigned to β. Differentiation of the equations $\bar{y}_h = \bar{y}_h(\bar{x})$ with respect to x_i yields

$$\frac{d\bar{y}_h}{dx_i} = \sum_{k=1}^{m} \frac{d\bar{y}_h}{d\bar{x}_k} \frac{d\bar{x}_k}{dx_i}.$$

Solving these equations for the derivative $d\bar{y}_h/d\bar{x}_k$, we get

$$\frac{d\bar{y}_h}{d\bar{x}_k} = \frac{N}{D},$$

where D is the functional determinant $\dfrac{d(\bar{x}_1, \cdots, \bar{x}_m)}{d(x_1, \cdots, x_m)}$ and N is the determinant obtained by replacing the element $d\bar{x}_k/dx_i$ in D by $d\bar{y}_h/dx_i$ for each i. Again using the fact that f and ϕ contain no linear terms and that $|dy/dx| < \alpha$, we see from differentiation of (3.1) that

$$\frac{d\bar{x}_k}{dx_i} = s_k\delta_{ki} + \frac{\partial f_k}{\partial x_i} + \sum_{j=1}^{n} \frac{\partial f_k}{\partial y_j}\,\frac{dy_j}{dx_i} = s_k\delta_{ki} + \xi_{ki}$$
$$\frac{d\bar{y}_h}{dx_i} = \sigma_h\,\frac{dy_h}{dx_i} + \frac{\partial \phi_h}{\partial x_i} + \sum_{j=1}^{n} \frac{\partial \phi_h}{\partial y_j}\,\frac{dy_j}{dx_i} = \sigma_h\,\frac{dy_h}{dx_i} + \eta_{hi}$$

where the ξ's and η's approach zero with $| x |$ uniformly as to C. We thus find

$$\frac{d\bar{y}_h}{d\bar{x}_k} = \frac{N}{D} = \frac{\sigma_h}{s_k}\,\frac{dy_h}{dx_k} + \theta_{hk},$$

where θ also tends uniformly to zero. Since, however, $| \sigma_h | \leqq \sigma < s \leqq | s_k |$, it follows that

$$\left| \frac{d\bar{y}_h}{d\bar{x}_k} \right| \leqq \left| \frac{\sigma_h}{s_k} \right| \cdot \alpha + | \theta_{hk} | < \frac{\sigma}{s}\,\alpha + | \theta_{hk} | < \alpha,$$

if β is sufficiently small. This completes the proof.

If $s > 1$, we may take $s^* = 1$ and we have the following corollary to Theorem 1:

COROLLARY. *If $s > 1$ we may always take β so small that any manifold C of $K(\alpha, \beta)$ is transformed by T and its iterates into manifolds of $K(\alpha, \beta)$.*

4. On the transformation of two manifolds. For the sake of brevity we shall hereafter, in reference to an expression a_i, depending on an index i, use the notation $\| a \| = \Sigma \, | a_i |$ where the summation is extended over the range of the index, which as a matter of fact will always be either $1, 2, \cdots, m$ or $1, 2, \cdots, n$.

Let C and C' be any two manifolds in $K(\alpha, \beta)$. If we consider two points with the same abscissas, one of them (x, y) on C and the other (x, y') on C', all the differences $y_h - y'_h$ will vanish for $x = 0$. Hence it is easily seen that the quantity $\| y - y' \| / \| x \|$ will have in a fixed neighborhood D of the origin a certain least upper bound $\omega \leqq 2\alpha n$. Now let $\bar{C}$ and $\bar{C}'$ denote the transforms of C and C' respectively. Denote by $\bar{\omega}$ the quantity analogous to ω determined from $\bar{C}$ and $\bar{C}'$. Then we may state as follows our

THEOREM 3. *The ratio $\bar{\omega}/\omega$ does not exceed a fixed number $\rho < 1$, if the domain D is chosen sufficiently small. For ρ we may take any real number greater than σ/s and less than 1. D may be chosen independently of C and C' if β is sufficiently small.*

Proof. Denote by Y and Y' the ordinates of C and C' respectively corresponding to the same abscissas X; $\bar{y}$ and $\bar{y}'$ the analogous ordinates of $\bar{C}$ and $\bar{C}'$. The points $(X, \bar{y})$ and $X, \bar{y}')$ will be the transforms of two points (x, y)

and (x', y') situated on C and C' respectively. From the hypothesis that $C \subset K(\alpha, \beta)$ we have

$$(4.1) \qquad |y| \leqq \max \left| \frac{dy}{dx} \right| \cdot \| x \| \leqq \alpha \| x \|, \text{ i.e. } \| y \| \leqq n\alpha \| x \|$$

and from the hypothesis that the f's and ϕ's begin with terms of the second degree we get

$$| X_i - s_i x_i | \leqq \eta(\| x \| + \| y \|) \leqq \eta(1 + n\alpha) \| x \|$$

where

$$\eta \left(\geqq \left| \frac{\partial f_i}{\partial x_j} \right|, \left| \frac{\partial f_i}{\partial y_j} \right|, \left| \frac{\partial \phi_i}{\partial x_j} \right|, \left| \frac{\partial \phi_i}{\partial y_j} \right| \right)$$

can be taken arbitrarily small by taking the dimensions of D sufficiently small. This inequality gives us

$$s | x_i | \leqq | s_i x_i | \leqq | s_i x_i - X_i | + | X_i | \leqq \eta(1 + n\alpha) \| x \| + | X_i |,$$

and hence

$$s \| x \| \leqq \eta m (1 + n\alpha) \| x \| + \| X \|$$

and thus we obtain

$$\| x \| \leqq \frac{\| X \|}{s - \eta m (1 + n\alpha)}.$$

If (x, y'_0) is the point of C' corresponding to the abscissas x, we have

$$(4.2) \qquad \| y - y'_0 \| \leqq \omega \| x \| \leqq \frac{\omega \| X \|}{s - \eta m (1 + n\alpha)}$$

$$(4.3) \qquad | y'_h - y'_{0h} | \leqq \max \left| \frac{dy'}{dx} \right| \cdot \| x' - x \| \leqq \alpha \| x' - x \|.$$

On the other hand by the very definition of X, x, x', y, y', we have

$$X_i = s_i x_i + f_i(x, y) \quad \text{and} \quad X_i = s_i x'_i + f_i(x', y');$$

and hence

$$s_i(x_i - x'_i) = f_i(x', y') - f_i(x, y).$$

Since the first derivatives of f_i and ϕ_i are all in absolute value less than η, we can write

$$s | x_i - x'_i | \leqq | s_i(x_i - x'_i) | \leqq \eta \| x - x' \| + \eta \| y - y' \|$$

$$s \| x - x' \| \leqq m\eta \| x - x' \| + m\eta \| y - y' \|$$

$$(4.4) \qquad \| x - x' \| \leqq \frac{m\eta \| y - y' \|}{s - m\eta}.$$

From (4.3) and (4.4) we obtain

$$| y'_h - y'_{0h} | \leqq \frac{\alpha m\eta \, \| y - y' \|}{s - m\eta}$$

$$| y'_h - y_h | \leqq | y'_h - y'_{0h} | + | y'_{0h} - y_h | \leqq \frac{\alpha m\eta \, \| y - y' \|}{s - m\eta} + | y'_{0h} - y_h |.$$

Therefore it follows (summing over h) that

$$\| y' - y \| \leqq \| y'_0 - y \| / \left(1 - \frac{\alpha n m\eta}{s - m\eta}\right) = \frac{(s - m\eta) \, \| y'_0 - y \|}{s - m\eta(1 + \alpha n)}.$$

From (4.2) we now obtain immediately

$$\| y' - y \| \leqq \frac{(s - m\eta)\,\omega}{[s - m\eta(1 + n\alpha)]^2} \, \| X \|. \tag{4.5}$$

By definition, we have $\bar{y}_h = \sigma_h y_h + \phi_h(x, y)$ and $\bar{y}'_h = \sigma_h y'_h + \phi_h(x', y')$. Hence

$$\begin{aligned} | \bar{y}_h - \bar{y}'_h | &\leqq | \bar{y}_h - \bar{y}'_h - \sigma_h(y_h - y'_h) | + | \sigma_h | \cdot | y_h - y'_h | \\ &= | \phi_h(x, y) - \phi_h(x', y') | + | \sigma_h | \cdot | y_h - y'_h | \\ &\leqq \eta(\| x - x' \| + \| y - y' \|) + | \sigma | \cdot | y_h - y'_h |. \end{aligned}$$

Hence

$$\| \bar{y} - \bar{y}' \| \leqq n\eta \, \| x - x' \| + (n\eta + \sigma) \, \| y - y' \|.$$

From (4.4) we obtain

$$\| \bar{y} - \bar{y}' \| \leqq \left[\frac{mn\eta^2}{s - m\eta} + n\eta + \sigma\right] \| y - y' \|,$$

while from (4.5) we see further that

$$\begin{aligned} \| \bar{y} - \bar{y}' \| &\leqq \left[\frac{mn\eta^2}{s - m\eta} + n\eta + \sigma\right] \frac{(s - m\eta)\,\omega}{[s - m\eta(1 + n\alpha)]^2} \, \| X \| \\ &= \left[\frac{\sigma}{s} + \xi(\eta)\right] \omega \, \| X \|, \end{aligned} \tag{4.6}$$

where $\xi(\eta)$ is real and tends to zero with η and therefore also with the dimensions of D. Since $\sigma < s$, we have $0 < \sigma/s + \xi(\eta) < \rho < 1$, provided that D is sufficiently small. Thus we have proved the desired result, namely, that $\bar{\omega}$, the least upper bound of $\| \bar{y} - \bar{y}' \| / \| X \|$ can not exceed $\rho\omega$.

The fact that D may be chosen independently of the particular manifolds C and C' considered, so long as they both belong to $K(\alpha, \beta)$ and β is first chosen sufficiently small, follows from two considerations: In the first place, according to Theorem 2, C, C', $\bar{C}$, $\bar{C}'$ have a common domain of definition independent of C and C', namely, the smaller of the two domains $| x | < \beta$ and $| x | < \beta s^*$. In the second place, the quantity $\xi(\eta)$ in (4.6) is independent of C and C'. As a matter of fact it is not difficult to see that by taking β

still smaller, if necessary, we may choose D to be the smaller of the two domains $|x| < \beta$ and $|x| < \beta s^*$. But this result is unnecessary for our purposes.

5. The invariant manifolds for $s > 1$. Let C_0 denote an arbitrary manifold in $K(\alpha,\beta)$ with the equations $y = y^{(0)}(x)$. We denote by C_k the k-th transform of C_0 under T. That is, C_0 is transformed into C_1 by T; C_1 is transformed into C_2 and so forth. In order to ensure that all the C_k also belong to $K(\alpha,\beta)$, we add the hypothesis concerning T that $s > 1$ and take β sufficiently small as described in the corollary of Theorem 2. Let D be the domain referred to in Theorem 3, dependent on the class $K(\alpha,\beta)$ but independent of the particular C and C' belonging to this class. It is necessarily contained in the region $|x| < \beta$. Let the equations of C_k be $y = y^{(k)}(x)$. We now prove our main result,

THEOREM 4. *There exists an invariant analytic manifold C^* with the equations $y = y^*(x)$, such that $\lim_{k=\infty} y^{(k)}(x) = y^*(x)$ uniformly in D.*

Proof. Let ω_k be the least upper bound of $\| y^{(k)} - y^{(k-1)} \| / \| x \|$ in D. By Theorem 3, we have $\omega_{k+1} \leqq \rho\omega_k$. It follows by induction that $\omega_k \leqq \rho^{k-1}\omega_1$. Thus we have

$$| y_i^{(k)}(x) - y_i^{(k-1)}(x) | \leqq \| y^{(k)}(x) - y^{(k-1)}(x) \| \leqq \omega_k \| x \| \leqq \rho^{k-1}\omega_1 m\beta,$$

and, since the series $\sum_{k=1}^{\infty} \rho^{k-1}\omega_1 m\beta$ is convergent, ρ being < 1, the uniform convergence of the sequence $y_i^{(0)}, y_i^{(1)}, y_i^{(2)}, \cdots$ follows from the classical theorem of Weierstrass on uniform convergence.

It follows from another theorem of Weierstrass[6] that C^* is analytic in the interior of D, since D is a domain of m complex dimensions or $2m$ real dimensions.

Furthermore C^* is invariant. For $\lim_{k=\infty} C_k = C^*$ while $\lim_{k=\infty} C_{k+1} =$ the transform of C^*, since T is continuous. It follows that C^* is identical with its transform.

THEOREM 5. *C^* is independent of the initially chosen manifold C_0.*

Proof. Let C'_k be another sequence of manifolds of the same type as C_k, i. e. each C'_k belongs to $K(\alpha,\beta)$; T carries C'_k into C'_{k+1}. Let $y = y^{(k)\prime}(x)$ be the equations of C'_k. Then, if ω is the least upper bound of

$$\| y^{(1)} - y^{(1)\prime} \| / \| x \|$$

in D, we find from Theorem 3 that

---

[6] Cf. W. F. Osgood, *Lehrbuch der Funktionentheorie*, vol. 2 (1924), first section, p. 13, 1. Satz.

$$| y_i^{(k)} - y_i^{(k)'} | \leqq \| y^{(k)} - y^{(k)'} \| \leqq \rho^k \omega \| x \| \leqq \rho^k \omega m \beta.$$

And, since $\rho < 1$, it follows that $\lim_{k=\infty} C'_k = \lim_{k=\infty} C_k$.

COROLLARY. *If C belongs to $K(\alpha, \beta)$ and is invariant it must necessarily coincide with the manifold C^* of Theorems 4 and 5 throughout D.*

Proof. Referring to the proof of Theorem 5, let us take $C'_0 = C$. Then, since C is assumed to be invariant, we have $C'_k = C$, and hence $C = \lim_{k=\infty} C_k = C^*$, throughout D where C^* is defined.

We might add the obvious remark that C^* can be continued analytically over the complete domain of definition of C and the two manifolds will then coincide throughout.

6. Note on real transformations. So far we have looked at the transformation T from the point of view of the complex domain. If, however, we are primarily interested in a real transformation and real analytic manifolds, we obtain similar results.

If T is a real transformation, we do not necessarily mean that all the coefficients in the power series appearing in (1.1) are real nor that real values for x and y represent real points. For, in order to reduce the linear terms of the transformation to the canonical form displayed in (1.1), it might be necessary to introduce conjugate imaginary variables, in which case the coefficients would not all be real and real values of the variables would not correspond to real points in the transformation. Since conjugate imaginary variables are associated with conjugate imaginary characteristic roots and since the characteristic roots s_i are all in absolute value greater than the characteristic roots σ_h, it is clear that, if any such conjugate imaginary pairs occur, none of the variables x can be paired with a variable y.

We briefly sketch the type of argument to be used in connection with the real theory. The C_0 of the previous paragraph may be taken in such a way that values of x corresponding to real points in the m-dimensional x-space will yield real points in the n-dimensional y-space. Then $C_1, C_2, \cdots$ will automatically have the same property, if T is real. The manifolds C_k are known to converge uniformly in a complex domain in which our real domain is imbedded. Hence our real manifolds will converge uniformly to a manifold which is necessarily both real and analytic.

It is also possible to use the method of this paper to treat directly real transformations which are not necessarily analytic, as, in fact, was done originally by Hadamard (*loc. cit.*) for the case $n = m = 1$. The invariant manifolds will then, of course, not necessarily be analytic.

CORNELL UNIVERSITY.

ON STRONG SUMMABILITY.*

By H. J. HAMILTON and J. D. HILL.

---

1. **Definitions.** Being given in the complex domain a matrix (a_{mk}) $(k, m = 1, 2, 3, \cdots)$, and a positive real number p we shall say (following Szász, *Brown University Lectures*, 1934-35) that a complex sequence $\{s_k\}$ is *strongly summable of order p to the complex number s by (a_{mk})*, or simply *strongly summable* to s, if for each $m > 0$ the series $\sigma_m \equiv \sum_k a_{mk} \mid s - s_k \mid^p$ is convergent, and if $\lim_m \sigma_m = 0$. If each convergent $\{s_k\}$ is strongly summable to its limit, then the matrix (a_{mk}) will be called *strongly regular*. The terms *summable* and *regular*, unmodified, are to be interpreted in the usual Silverman-Toeplitz sense.

We shall say that a matrix is *of type U* if it has the property that whenever any sequence is strongly summable to both s and t, we are permitted to conclude that $s = t$. The phrase *of type U for convergent sequences* will mean that the strong summability of any *convergent* sequence to both s and t implies $s = t$; and analogously for bounded sequences.

2. **Strong regularity.** To find necessary and sufficient conditions for strong regularity it is convenient to introduce three lemmas. We mention in passing that the index p plays no essential rôle here.

LEMMA 1. *In order that (a_{mk}) be strongly regular it is necessary and sufficient that it be regular for real non-negative null sequences.*

LEMMA 2. *In order that (a_{mk}) be regular for real non-negative null sequences it is necessary and sufficient that it be regular in the space (c_0) of real null sequences.*

LEMMA 3. *In order that (a_{mk}) be regular in (c_0) it is necessary and sufficient that*

$$\lim_m a_{mk} = 0 \qquad (k = 1, 2, \cdots), \tag{1}$$

$$\sup_m \sum_k \mid a_{mk} \mid < \infty. \tag{2}$$

---

* Received January 24, 1938; Revised April 25, 1938.

Proofs. The first is obvious, and the second follows from the fact that an arbitrary real null sequence may be written as the difference of two real non-negative null sequences. To prove the third observe first that in case (a_{mk}) is real, a slight modification of the regularity proof for convergent sequences[1] suffices to establish (1) and (2) as necessary and sufficient.

In the general case we may write $a_{mk} = a'_{mk} + ia''_{mk}$ where a'_{mk}, a''_{mk} are real. It is clear that (a_{mk}) is regular in (c_0) if and only if each of the matrices (a'_{mk}), (a''_{mk}) has the same property. In order that the latter be true one easily shows, by the above statement for real matrices, that (1) and (2) are necessary and sufficient, and the proof is complete.

From these three lemmas we obtain the following theorem, in answer to a question raised by Dr. Szász.

THEOREM 1. *In order that* (a_{mk}) *be strongly regular the conditions* (1) *and* (2) *are necessary and sufficient.*

Since the Silverman-Toeplitz conditions include (1) and (2) we observe that every regular matrix is strongly regular.

3. Matrices of type *U*. To treat the question of uniqueness of the strong limit we need the following lemma.

LEMMA 4. *For each* $p > 0$ *there exists a constant* K, *depending on* p, *such that for all values of* a *and* b *we have* $|a + b|^p \leqq K(|a|^p + |b|^p)$.

Proof. We may assume that $a + b \neq 0$ and hence that a, say, is not zero. Then we have

$$\frac{|a+b|^p}{|a|^p + |b|^p} \leqq \frac{(|a| + |b|)^p}{|a|^p + |b|^p} = \frac{(1 + |b/a|)^p}{1 + |b/a|^p}.$$

Now it is easy to verify that if $0 < p \leqq 1$, $F(x) \equiv (1 + x)^p/(1 + x^p) \leqq 1$ for $0 \leqq x < \infty$, and if $p > 1$, $F(x) \leqq 2^{p-1}$ for $0 \leqq x < \infty$. Consequently, the result follows with K equal to 1 in the first case and 2^{p-1} in the second.

It will be useful to investigate first of all some of the properties of matrices for which a constant $M \geqq 0$ exists such that

(3) $$a_{mk} \geqq 0 \text{ for each } m > 0 \text{ and all } k > M,$$

(4) $$\lim_{m} a_{mk} = 0 \text{ for } 1 \leqq k \leqq M.$$

---

[1] See, for example, Banach, *Théorie des opérations linéaires*, pp. 90-91.

For such matrices the strong limit has the property of finite additivity as shown by the following theorem.

THEOREM 2. *If* $\{s_k\}$ *and* $\{t_k\}$ *are strongly summable to* s *and* t, *respectively, by a matrix* (a_{mk}) *which satisfies* (3) *and* (4), *then for all values of* a *and* b *the sequence* $\{as_k + bt_k\}$ *is strongly summable to* $as + bt$.

Proof. Using Lemma 4 we may establish the inequality

$$\left| \sum_k a_{mk} \left| (as_k + bt_k) - (as + bt) \right|^p \right| \leqq \sum_{k \leqq M} |a_{mk}| \; |a(s_k - s) + b(t_k - t)|^p$$
$$+ K \, |a|^p \sum_{k > M} a_{mk} \, |s_k - s|^p + K \, |b|^p \sum_{k > M} a_{mk} \, |t_k - t|^p$$

for each $m > 0$, and the conclusion is apparent.

THEOREM 3. *If* (a_{mk}) *satisfies* (3), (4), *and if*

(5) $\sum_k a_{mk}$ *is convergent for each* $m > 0$,

then the condition

(6) $$\overline{\lim_m} \left| \sum_k a_{mk} \right| > 0$$

is necessary and sufficient that (a_{mk}) *be of type* U.

Proof. If (a_{mk}) is of type U the sequence $1, 1, 1, \cdots$, which is strongly summable to 1, can not be strongly summable to 0; hence the necessity. On the other hand, if $\{s_k\}$ is strongly summable to s and also to t, then by Theorem 2 the sequence $\{(1)s_k + (-1)s_k\} = 0, 0, 0, \cdots$ is strongly summable to $s - t$. Thus we must have $\lim_m |s - t|^p \sum_k a_{mk} = 0$, and if (6) holds this is possible only if $s = t$.

THEOREM 4. *If* $\{s_k\}$ *is strongly summable to* s *by a matrix* (a_{mk}) *which satisfies* (3), (4), (5), (6), *then* s *must be a limit point of* $\{s_k\}$.

Proof. If we set $\sigma'_m \equiv \sum_{k \leqq M} a_{mk} \, |s_k - s|^p$, $\sigma''_m \equiv \sum_{k > M} a_{mk} \, |s_k - s|^p$ then $\sigma'_m + \sigma''_m \to 0$, whence, by (4), $\sigma''_m \to 0$. But the latter is impossible if s is not a limit point of $\{s_k\}$. For in that event a number $d > 0$ exists such that $|s_k - s|^p > d$ for all k. Then $\sigma''_m > d \sum_{k > M} a_{mk}$, and since by (4) and (6), $\overline{\lim_m} \left| \sum_k a_{mk} \right| = \overline{\lim_m} \sum_{k > M} a_{mk} > 0$, we have $\overline{\lim_m} \, \sigma''_m > 0$.

The foregoing property of strong summability may be used in obtaining the following comparison with ordinary summability.

THEOREM 5. *Suppose that* (a_{mk}) *satisfies* (3), (4), *and*

$$\lim_{m} \sum_{k} a_{mk} = 1. \tag{7}$$

Then if $\{s_k\}$ *is strongly summable of order 1 to s by* (a_{mk}), *it is also summable to s by the same matrix. Moreover, the converse is false for every order p.*

Proof. From the inequality

$$\left| \sum_{k} a_{mk}(s - s_k) \right| \leqq \sum_{k \leqq M} | a_{mk} | \; | s - s_k | + \sum_{k > M} a_{mk} \, | s - s_k |$$

we see that $\sum_{k} a_{mk}(s - s_k) = o(1)$. Consequently we obtain

$$\sum_{k} a_{mk} s_k = s \sum_{k} a_{mk} + o(1),$$

and the first part of the conclusion follows.

Since the sequence $\{(-1)^{k-1}\}$ is $(C, 1)$ summable to 0, the falsity of the converse is apparent from Theorem 4.

Proceeding to the general case we begin by establishing the following result.

THEOREM 6. *In order that a strongly regular matrix* (a_{mk}) *be of type U for convergent sequences, the condition* (6) *is necessary and sufficient.*

Proof. The necessity follows as in the proof of Theorem 3. To establish the sufficiency let $\{s_k\}$ be an arbitrary convergent sequence. By the assumed regularity we have

$$\lim_{m} \sum_{k} a_{mk} \, | s - s_k |^p = 0, \tag{a}$$

where $s \equiv \lim_{k} s_k$. Denote by t any number for which

$$\lim_{m} \sum_{k} a_{mk} \, | t - s_k |^p = 0. \tag{b}$$

Setting $u_k \equiv | s - s_k |^p + | t - s_k |^p$ we obtain

$$\lim_{k} u_k = | t - s |^p, \tag{c}$$

and from (a), (b),

$$\lim_{m} \sum_{k} a_{mk} u_k = 0. \tag{d}$$

Now from (6) we infer the existence of a sequence $\{m_n\}$ tending to infinity, and of a number $\lambda \neq 0$, such that

$$\lim_{n} \sum_{k} \lambda a_{m_n k} = 1. \tag{e}$$

Since the matrix $(b_{nk}) \equiv (\lambda a_{m_n k})$ satisfies (1), (2), (e), it is regular in the Silverman-Toeplitz sense. Consequently, from (c) and (d) we find that $\lim_n \sum_k b_{nk} u_k = | t - s |^p = 0$, whence $t = s$.

At this point the question naturally arises whether or not conditions (1), (2), (6) are sufficient to define the strong limit uniquely in general, and the following example shows that they are insufficient even for the case of real matrices and bounded sequences.

Example 1. Define $a_{m,3n-2} = 1/m$ $(m \geqq 3n - 2)$, $a_{m,3n-1} = 1/m$ $(m \geqq 3n - 1)$, $a_{m,3n} = -1/m$ $(m \geqq 3n)$ for $n = 1, 2, 3, \cdots$, and let $a_{mk} = 0$ otherwise. Then (a_{mk}) is strongly regular, and condition (6) is satisfied. But if we let $s_{3n-2} = 1$, $s_{3n-1} = 0$, $s_{3n} = -1$ for $n = 1, 2, 3, \cdots$, one easily shows that the sequence $\{s_k\}$ is strongly summable by (a_{mk}) to both 0 and 2.

It may be of interest to point out that this example suffices also to show that the conclusion of Theorem 4 is no longer valid in the absence from the hypotheses of condition (3).

In view of Example 1 it becomes necessary to seek further conditions on (a_{mk}) in order to insure uniqueness for strongly summable sequences that are not necessarily convergent. The problem of finding such conditions which are necessary as well as sufficient remains unsolved. Certain sufficient conditions, however, are introduced in the following lemmas and theorems.

LEMMA 5. *If, for any subsequence* $\{m_n\}$ *of* $\{m\}$, $(a_{m_n k})$ *is of type* U, *then* (a_{mk}) *is of type* U.

LEMMA 6. *If, for any bounded sequence of complex numbers* $\{c_m\}$, $(c_m a_{mk})$ *is of type* U, *then* (a_{mk}) *is of type* U.

LEMMA 7.[2] *If either* (a'_{mk}) *or* (a''_{mk}) *is of type* U, *then* (a_{mk}) *is of type* U.

The proofs of these lemmas are immediate consequences of the definition of matrices of type U.

THEOREM 7. *In order that a given matrix* (a_{mk}) *be of type* U *in general, it is sufficient that conditions* (4), (5), (6), *and the following condition* (8) *should be satisfied.*

---

[2] The authors are indebted to Dr. Otto Szász for suggesting the introduction of Lemma 7, and also for pointing out the generalization of Theorems 7 and 8 obtained by replacing the condition $\gamma = \pi/4$ of the original manuscript by the condition $\gamma < \pi/2$.

(8) *There exist a value* ϕ_{mk} *of* $\arg a_{mk}$ $(m, k = 1, 2, 3, \cdots)$, *a sequence* $\{\phi_m\}$, *a positive constant* $\gamma < \pi/2$, *and a constant* $M \geqq 0$, *such that, for each* $m > 0$ *and for all* $k > M$ *we have* $\phi_m - \gamma \leqq \phi_{mk} \leqq \phi_m + \gamma$.

Remark. Stated in geometric language, condition (8) postulates the existence in the complex plane of congruent sectors Q_m of angle less than π, such that for each m $\arg a_{mk}$ lies in Q_m for all $k > M$.

Proof. Setting $e^{-i\phi_m}a_{mk} \equiv \alpha_{mk} + i\beta_{mk}$, where α_{mk} and β_{mk} are real, we see by Lemmas 7 and 6 that it will suffice to prove that (α_{mk}) is of type U. Now $\alpha_{mk} = | a_{mk} | \cos(\phi_{mk} - \phi_m) \geqq | a_{mk} | \cos \gamma \geqq 0$ for each $m > 0$ and all $k > M$, so that (α_{mk}) satisfies (3). That (α_{mk}) satisfies (4) and (5) is clear. Moreover, from the inequality $| \beta_{mk} | \leqq \alpha_{mk} \tan \gamma$ $(m > 0,\ k > M)$, it follows that

$$\Big| \sum_{k>M} a_{mk} \Big| \equiv \Big| \sum_{k>M} e^{-i\phi_m} a_{mk} \Big| \leqq \sum_{k>M} \alpha_{mk} + \sum_{k>M} | \beta_{mk} | \leqq (1 + \tan \gamma) \sum_{k>M} \alpha_{mk}$$

for all m, whence, in view of (4) and (6), (α_{mk}) satisfies (6). The conclusion now follows from Theorem 3.

The question of whether any such "sector condition" as (8) is necessary for uniqueness is answered in the negative by the next example.

Example 2. Define $a_{mm} = a_{m,m-2} = 1$, $a_{m,m-1} = -1$ for $m \geqq 3$, and $a_{mk} = 0$ otherwise. Then (a_{mk}) satisfies (1), (2), and (6), but not (8). It is nevertheless of type U. For supposing that $\{s_k\}$ is strongly summable to s, we have, for $m \geqq 3$,

$$\sigma_m \equiv \sum_k a_{mk} | s - s_k |^p = | s - s_{m-2} |^p - | s - s_{m-1} |^p + | s - s_m |^p \to 0,$$

whence

$$\sigma_m + \sigma_{m+1} + \sigma_{m+2} = | s - s_{m-2} |^p + | s - s_m |^p + | s - s_{m+2} |^p \to 0,$$

so that $s_m \to s$. Thus (a_{mk}) sums convergent sequences only, and, indeed, only to their proper limits.

THEOREM 8. *In order that a given matrix* (a_{mk}) *be of type* U *for bounded sequences, it is sufficient that conditions* (5), (6), *and the following condition* (9) *should be satisfied.*

(9) *There exist a value* ϕ_{mk} *of* $\arg a_{mk}$ $(m, k = 1, 2, 3, \cdots)$, *sequences* $\{\phi_m\}$, $\{k_n{}^m\}$, *and a positive constant* $\gamma < \pi/2$ *such that* (i) *for each* $m > 0$ *and for all* $k \neq k_n{}^m$ *we have* $\phi_m - \gamma \leqq \phi_{mk} \leqq \phi_m + \gamma$, *and* (ii) $\lim_m \sum_n | a_{mk_n{}^m} | = 0$.

Remark. It is clear that any matrix which satisfies (4) and (8) also satisfies (9), whereas the converse is not true.

Proof. For each $m > 0$ define b_{mk} as 0 if $k = k_n^m$, and as a_{mk} if $k \neq k_n^m$ $(k, n = 1, 2, 3, \cdots)$. It is clear that it will suffice to prove that (b_{mk}) is of type U; but this is immediate, since (b_{mk}) satisfies the conditions of Theorem 7, with $M = 0$.

Theorems 7 and 8 can be extended by application of Lemma 5 in an obvious manner. (It may be noted that the condition $\varliminf_m |\sum_n a_{mk}| > 0$ insures the satisfaction of (6) for all subsequences $\{m_n\}$—and of course conversely.) That no such "sector condition" as (8) or (9) is necessary for uniqueness of the strong limit in any case, is evident from Example 2 above.

POMONA COLLEGE,
MICHIGAN STATE COLLEGE.

INVARIANTS DERIVED FROM LOOPING COEFFICIENTS.*

By E. R. van Kampen.

A looping coefficient[1] was first introduced by Brouwer for two non-intersecting manifolds of complementary dimensions k and $n-k-1$ embedded in n-dimensional space. Let the product of the two manifolds be transformed into the unit sphere of n-space by means of the vectors from points of the first manifold to points of the second manifold. Then Brouwer defines the looping coefficient of the two manifolds as the degree of the resulting transformation. It is essential for this definition that the manifolds are embedded in ordinary n-space, but if the manifolds are replaced by cycles of the same dimensions, the definition retains its meaning. When the identity of the looping coefficient of two non-interesting cycles of complementary dimensions with the intersection number of one of the cycles and any chain bounded by the other cycle was established, the way was paved for the introduction of a looping coefficient for cycles which are homologous to zero in any arbitrary orientable n-dimensional manifold M_n. At this stage the looping coefficients still were characteristics of the cycles only and did not contain any invariant property of the manifold M_n itself.

The next step was the introduction of looping coefficients also for so-called zero divisors of M_n, that means for cycles in M_n of which a certain integral multiple is homologous to zero in M_n. The definition was now formulated as follows: If α times the cycle a_1 is the boundary of the chain c and the cycle a_2 of complementary dimension to a_1, is a zero-divisor and does not meet a_2, then the looping coefficient $L(a_1, a_2)$ is the intersection number of c and a_2 divided by α. The condition that a_2 is a zero-divisor is necessary and sufficient for $L(a_1, a_2)$ to be independent of the particular choice of c. In this last form the looping coefficients are a priori capable of carrying invariant properties of the manifold M_n, since it turns out that $L(a_1, a_2)$ is determined modulo 1 by the homology classes of the cycles a_1 and a_2. As a consequence of (6) (cf. **1**), the looping coefficients of zero-divisors do not carry

---

* Received March 15, 1938; Revised April 11, 1938.

[1] L. E. J. Brouwer, "On looping coefficients," *Proceedings of the Royal Academy in Amsterdam*, vol. 15 (1913), pp. 113-122. Compare also: G. de Rham, "Sur l'analysis situs des variétés à n dimensions," *Journal de Mathématique*, vol. 10 (1931), pp. 115-200; in particular pp. 156-165; H. Seifert, "Verschlingungsinvarianten," *Sitzungsberichte der Preussischen Akademie, Physik.-Math. Klasse*, 1933, pp. 811-828.

invariant properties of M_n if the complementary dimensions k and $n-k-1$ are distinct, since in that case any base may be used for the torsion group (the group of zero-divisors) both for the dimension k and the dimension $n-k-1$.

If $n=2k+1$, however, the complementary dimensions are equal and the following problem arises. Let two manifolds M_n and M'_n have simply isomorphic groups of k-dimensional torsion. In which case is it possible to conclude from their looping coefficients alone that M_n and M'_n are not homeomorphic. This problem may be solved by transforming the matrix $L(a_i, a_j)$ into one of a certain collection of non-equivalent normal forms, by a suitable choice of the base a_i for the group of k-dimensional torsion. It is easily seen [2] that in considering this problem one may consider separately for each given prime number p, the zero-divisors of which the order is a power of p.

For even k, the problem was solved by G. de Rham.[3] It turns out that no new invariant of the manifold may be obtained in this case by the consideration of zero-divisors of which the order is a power of an odd prime number p, while for the zero-divisors of order 2^k certain unexpected a priori possibilities arise for the matrix $L(a_i, a_j)$. However thus far no examples are known of manifolds for which the matrix $L(a_i, a_j)$ assumes such an anomalous form.

For odd k and $p \neq 2$, the problem was solved by H. Seifert.[4] In this note the case of odd k will be considered without the restriction that $p \neq 2$. Certain of the considerations of Seifert will be repeated (cf. **2-5**). This is due partly to certain modifications necessitated by the inclusion of the case $p=2$ and partly to the desire to treat the cases $p \neq 2$ and $p=2$ in the same way. In contrast to the former case, the main difficulty of the latter case is the consideration of the influence on each other of zero-divisors of which the orders are adjoining powers of 2 (cf. **11**).

In **12** the normal forms arrived at are computed for an infinite collection of known 3-dimensional manifolds. These examples clearly indicate how little the homology and intersection theory can contribute to the fundamental problems concerning manifolds even of such a low dimension.

1. The looping coefficient $L(a, b)$ of a k-dimensional zero-divisor a of order α and a $(n-k-1)$-dimensional zero-divisor b of order β in an orientable manifold M_n is given by

$$L(a, b) = S(c, b)/\alpha, \tag{1}$$

---

[2] H. Seifert, *loc. cit.*, pp. 816-817.

[3] *Loc. cit.*, pp. 159-165.

[4] *Loc. cit.*, p. 823.

where c has αa as boundary. Here $S(c, b)$ is the intersection number and a and c do not have a point in common. The following properties are well known:

(2) $L(a, b)$ is uniquely determined modulo 1 if only the homology classes of a and b are given.

(3) $L(a, b) \equiv (-1)^{kn+1} L(b, a)$ modulo 1.

(4) $L(\alpha_1 a_1 + \alpha_2 a_2, b) = \alpha_1 L(a_1, b + \alpha_2 L(a_2, b)$.

(5) If $\alpha a \sim 0$ and $\beta b \sim 0$, then $L(a, b)$ may be written as a fraction with the greatest common divisor (α, β) of α and β as denominator. In particular if α and β are relative prime, then $L(a, b)$ is an integer.

(6) If a_i, $i = 1, 2, \cdots$ are the k-dimensional torsion numbers of M_n, then they are also the $(n - k - 1)$-dimensional torsion numbers of M_n. Furthermore if a_i is a base of the group of k-dimensional zero divisors, such that $a_i a_i \sim 0$, then a base b_i may be found for the $(n - k - 1)$-dimensional zero divisors in such a way that $\alpha_i b_i \sim 0$ and $\alpha_i L(a_i, b_j) = \delta_{ij}$. Here δ_{ij} is the Kronecker symbol.

(7) A statement analogous to (6) remains true for those k-dimensional zero-divisors of which the order is a power of a fixed prime number p.

This statement follows easily from the last remark in (5). The same remark shows also that the problem under consideration is solved in the general case if it has been solved for the zero-divisors of prime power order corresponding to every prime divisor of the torsion numbers of M_n.

2. From now on, k is an odd integer, $n = 2k + 1$ and M_n is a fixed orientable n-dimensional manifold. For a fixed prime number p, let G denote the group formed by these k-dimensional zero divisors of G, of which the order is a power of p. Let $a_1, \cdots, a_2$ be a base of G and let every relation between the a_i be a consequence of the order relations

$$p^{e_i} a_i \sim 0, \qquad 0 \leqq e_1 \leqq e_2 \leqq \cdots \leqq e_s. \tag{8}$$

Let A be the matrix with elements $g_{ij} = p^{e_i} L(a_i, a_j)$; then (2) shows that the integers g_{ij} are uniquely determined modulo p^{e_i}, while (3) shows that

$$p^{e_j} g_{ij} \equiv p^{e_i} g_{ji} \text{ modulo } p^{e_i + e_j}. \tag{9}$$

Since the transformation from one base of G to another base of G may be made by means of a non-singular matrix modulo p, it is clear that (7) is equivalent to the statement that A is non-singular modulo p. Since, by (9),

g_{ij} is divisible by $p^{e_i - e_j}$, whenever $e_i > e_j$, the symmetric submatrices $A^{(e)}$ of A, which correspond to sets of subscripts $l, l+1, \cdots, m$, where

$$(10) \qquad e = e_l = e_{l+1} = \cdots = e_m, \quad e_{l-1} < e, \quad e_m < e_{m+1},$$

are also non-singular modulo p.[5] Now $A^{(e)}$, being non-singular modulo p, has an inverse matrix modulo p^e which also has integer elements. This inverse may be obtained by multiplying the ordinary inverse by an appropriate integer prime to p. If $e^i > e^m = e$ then the sum of a_i and any linear combination of the generators $a_l, a_{l+1}, \cdots, a_m$ may be used as generator instead of a_i. With the help of the inverse matrix of $A^{(e)}$ one may easily find linear combinations in such a way that for the new matrix A,

$$g_{ji} = g_{ij} = 0, \quad j = l, l+1, \cdots, m, \quad i > m.$$

Thus A may be reduced to diagonal form, each element of the diagonal being one of the matrices $A^{(e)}$.

If this reduction has been completed, let H be the subgroup of G generated by the elements $a_l = b_1$, $a_{l+1} = b_2, \cdots, a_m = b_r$ of G, $r = m - l + 1$. Then any relation between the b_i is a consequence of the order relations $p^e b_i \sim 0$. Let $[c_i, c_j]$ denote, for any pair of elements c_i, c_j of H, the number $P^e L(c_i, c_j)$. Then, by (1), (2), and (4) the $[c_i, c_j]$ are integers uniquely determined modulo p^e and the $[c_i, c_j]$ are linear in c_i. Since k and n are odd, (3) implies that $[c_i, c_j] = [c_j, c_i]$.

Furthermore, the matrix

$$(11) \qquad B = \| h_{ij} \| = \| [b_i, b_j] \| = \| [b_j, b_i] \| = \| h_{ji} \|$$

is non-singular modulo p, since it is identical with $A^{(e)}$.

The next problem is to determine a number of non-equivalent normal forms into one of which every such matrix B may be transformed by a change of base of H. If the new generators of H are given in terms of the old generators by left multiplication with a matrix P, which is necessarily non-singular modulo p, then the matrix B is replaced by the matrix PBP'. Thus the determinant of B may be multiplied by any quadratic residue modulo p^e but not by any other factor.

In what follows a matrix B will occasionally be considered as belonging to two or more groups H of which the generators have two or more distinct orders p^e. Wherever necessary the exponent e will be specified by placing modulo p^e after a statement on B.

3. Let B be the matrix (11) corresponding to a group H which is generated by r independent elements $b_1, \cdots, b_r$, each of order p^e.

---

[5] H. Seifert, *loc. cit.*, pp. 818-819.

First it will be shown[6] that B *may be transformed into diagonal form if* p *is odd.* In fact, if h_i is not divisible by p, one has $[b_i, mb_i] = mh_{ii} \equiv 1$. modulo p^e for a suitably chosen integer m. If one replaces b_j by the new generator $b_j - mb_ih_{ij}$, the element h_{ij} will be replaced by zero. Proceeding in this fashion one may replace all elements in the i-th row and i-th column (except h_{ii} itself) by zero for any i, for which $h_{ii} \not\equiv 0$ modulo p. Now suppose that all reductions of this type have been made and that h_{ii} is divisible by p. Since B is non-singular modulo p, it follows that at least one element in the i-th row, say h_{ij}, is not divisible by p. Since all reductions of the preceding type have been made, h_{jj} must be divisible by p. If one replaces b_i by the new generator $b_i + b_j$, the i-th diagonal element in the new matrix B is $h_{ii} + h_{jj} + 2h_{ij}$. Thus if $p \neq 2$, the new diagonal element is not divisible by p and the preceding reduction may be continued until B has been transformed into diagonal form.

4. Suppose that $p \neq 2$ and that B has been transformed into diagonal form. If u_1 and u_2 are arbitrary diagonal elements of B, corresponding to the generators b_1 and b_2 of H, then it is possible to replace u_1 by 1, changing u_2 if necessary, but leaving all other elements of B unchanged.[7] It will first be shown that the subgroup of H generated by the elements b_1 and b_2 contains an element $c = \alpha_1 b_1 + \alpha_2 b_2$ such that

$$[c, c] = \alpha_1^2 u_1 + \alpha_2^2 u_2 \equiv 1 \text{ modulo } p^e. \tag{12}$$

If u_1 (or u_2) is a quadratic residue of p^e, one may choose α_2 (or α_1) equal to zero and determine α_1 (or α_2) in such a way that (12) is satisfied. If u_1 and u_2 are not quadratic residues of p^e, let v be a quadratic non-residue of p^e such that for some r,

$$1 + r^2 \equiv v \text{ modulo } p^e. \tag{13}$$

One may take as v the smallest positive quadratic non-residue modulo p. For any quadratic residue (non-residue) modulo p is also a quadratic residue (non-residue) modulo p^e. The group of prime residual classes modulo p^e is cyclic, so that $u_1 v$ and $u_2 v$ are quadratic residues. Thus the integers a_1 and a_2 may be found in such a way that

$$\alpha_1^2 u_1 v \equiv 1, \qquad \alpha_2^2 u_2 v \equiv r^2 \text{ modulo } p^e.$$

Equation (12) is now an immediate consequence of (13).

---

[6] H. Seifert, *loc. cit.*, p. 822.

[7] Compare H. Seifert, *loc. cit.*, p. 823, footnote. It seems that the elaboration in 4 is necessary to complete the indications given there.

The element c may obviously be used as one of the generators of the subgroup of H generated by b_1 and b_2. By (12) the corresponding diagonal element of the new matrix B is 1. If the other generator of the subgroup H is chosen according to the method of **3**, and the remaining generators of H are left unchanged, the object stated at the beginning of **4** is obviously reached.

Clearly one may use the above method to replace the diagonal elements of B, except perhaps one, by the number 1. The remaining element different from 1 may be placed in the position h_{11} and may be multiplied with an arbitrary factor α^2 prime to p^e by multiplying the corresponding generator with α. Thus it is clear that *a matrix* B *corresponding to a triple of numbers* $p \neq 2, e, r$ *may be transformed into the unit matrix* E_r, *if* $\det B$ *is a quadratic residue modulo* p^e, *and* B *may be transformed into the matrix* $E_r^{(p)}$, *obtained from the unit matrix* E_r *by changing the first diagonal element into an arbitrary quadratic non-residue modulo* p^e, *if* $\det B$ *is not a quadratic residue modulo* p^e. *Moreover, if* $p \neq 2$, *then two matrices, which are non-singular modulo* p, *are equivalent modulo* p^e, *if and only if they are equivalent modulo* p.

5. The structure of the group H is determined by the numbers p, e, r, which in turn are determined by (10) in terms of the fundamental relations (8) of G. However, it is clear that the subgroup H of G is not only not uniquely determined by G but even not uniquely determined by the matrix A of the numbers g_{ij}. In fact let A be reduced, as described in **2**, to a diagonal form, each diagonal element of which is one of the matrices $A^{(e)}$. If i and j are such that $e_i > e_j$, add $p^{e_i - e_j} a_i$ to the generator a_j of G. Apply the reduction of **2** to the mattrix A in its new form. The j-th generator remains equal to $a_j + p^{e_i - e_j} a_i$, so that the group H, belonging to $e = e_j$, has been replaced by a different group. Accordingly the considerations of **3** and **4**, for the case $p \neq 2$, must be completed by an investigation of the changes of $\det A^{(e)}$, when the group H is replaced by a different group. This is very easily done by means of the following statement.

If two different matrices B and C both are obtained from the same group G as the matrix $A^{(e)}$ corresponding to two different subgroups H and H', both of which have r generators of order p^e, then they satisfy a relation of the form

$$C \equiv PBP' \text{ modulo } p^c \tag{14}$$

for some matrix P which is non-singular modulo P. Here c is the least positive value of $| e_i - e |$ and the e_i are the different orders of generators of G.

In fact the generators b'_i of H' may be expressed in terms of the generators b_i of H, $i = 1, \cdots, r$ and the remaining elements a_j of a set of generators of G (which includes the b_i) in the form $b' = Pb + Qa$, where b', b and a are vectors with components b'_i, b_i and a_j. It is easily verified that P must be a non-singular square matrix modulo p, so that new generators of H are defined by $b'' = Pb$. The matrix $A^{(e)}$ which corresponds to the new generators b'' is clearly $D = PBP'$.

Thus in order to prove (14) it is sufficient to show that if the group H is modified by the addition of linear combinations of generators of order $\neq p^e$ to the generators b'' of order p^e, then the congruence classes modulo p^e of the corresponding matrices $A^{(e)}$ are the same.

If a_j has the order $p^{e_j} < p^e$ and if the generator b''_i is replaced by $b'_i = b''_i + \alpha a_j$, then one has for the i-th diagonal element of $A^{(e)}$,

$$p^e L(b'_i, b'_i) = p^e L(b''_i, b''_i) + p^{e-e_j}[2\alpha p^{e_j} L(a_j, a_j) + \alpha^2 p^{e_j} L(a_j, a_j)]$$

and similar changes take place in the other elements of B. If a_j has the order $p^{e_j} > p^e$, then only linear combinations of the form $\alpha p^{e_j - e} a_j$ may be added to the generators of H, and a similar change in the elements of $A^{(e)}$ results. Since passage from b''_i to b'_i is possible by a finite number of steps of the two types described it clearly follows that (14) holds.

As a direct consequence of (14) and the last statement in **4** one obtains the following final statement for the case under consideration: *If* $p \neq 2$, *the invariants of the different matrices* $A^{(e)}$ *are also invariants of* A *itself.*

6. Let B be the matrix (11) which corresponds to a group H with r generators each of order 2^e. *Then* B *may be transformed into diagonal form, if and only if at least one diagonal element of* B *is odd.* Since B is non-singular the necessity of this condition is made evident by the remark that, if every diagonal element of B is even, then $[a, a]$ is even for every a in H.

For the proof of the sufficiency let it be remarked first that, if no diagonal element of B is odd, then the number of rows of B is even. Otherwise B would be singular modulo 2, since every term of B would either contain a diagonal element or have a symmetric term distinct from itself. It also follows that the coefficient of any diagonal element of B in $\det B$ is even, so that B remains non-singular modulo 2^e if any number is added to one of its diagonal elements.

If B has at least one odd diagonal element, B may be reduced by the method of **4** to the form

$$B = \| h_{ij} \| = \left(\begin{array}{cccc} h_1 & & & \\ & \cdot & & \\ & & h_s & \\ & & & B'' \end{array} \right)$$

where B'' is a square matrix without odd diagonal element and the empty spaces represent terms which are equal to zero. If one replaces the generator b_{s+1} by $b_{s+1} + b_1$, then $h_{1\,s+1}$, $h_{s+1\,1}$ and $h_{s+1,s+1}$ increase by h_1, so that the new $(s+1)$-th diagonal element is an odd number. Since the matrix, which replaces B'', is non-singular by the above remark, one can reduce the elements $h_{1\,s+1}$ and $h_{s+1\,1}$ to zero by adding a suitable linear combination of $b_{s+1}, \cdots, b_r$ to b_1. The first diagonal element in the resulting matrix B is again odd, since it is the only element $\neq 0$ in its column. Thus the number of odd diagonal elements of B has been increased by one. This procedure, combined with the method in **3**, reduces B to diagonal form.

7. Let B be the matrix (11) belonging to a group H with r generators of order 2^e and suppose that B has diagonal form. If $e = 1$, B has the form E_r modulo p^e, so that no further reduction is necessary. If $e = 2$, the diagonal elements of B may be taken equal to $+1$ or -1. It will be shown that *a matrix B corresponding to a triple of numbers $p = 2$, $e = 2$, r may be transformed into one of the non-equivalent forms E_{rm}, $0 \leqq m < 3$, $m \leqq r$, where the first m elements in the diagonal of E_{rm} are -1, while the remaining elements are $+1$*, (in particular $E_{r0} = E_r$). First it is easily seen that any four elements equal to -1 in the diagonal of B may be replaced simultaneously by $+1$. In fact if $[a_i, a_i] = -1$, $i = 1, 2, 3, 4$, then $[b_i, b_i] \equiv +1$ modulo 4, where

$$b_1 = a_1 + a_2 + a_3,\; b_2 = a_1 - a_2 + a_4,\; b_3 = -a_1 + a_3 + a_4,\; b_4 = a_2 - a_3 + a_4.$$

Thus if the b_i replace the a_i as generators of H, four elements -1 in the diagonal of B are replaced by -1. This proves that any of the matrices B under consideration may be transformed into one of the above normal forms E_{rm} and also that *m is determined as the residual class modulo 4 of the number of diagonal elements -1 in B*. It remains to prove that the above normal forms E_{rm} are not equivalent. The value of $\det B$ may be used to prove that E_{rm} and E_{rn} are not equivalent if $m - n$ is odd. The non-equivalence of E_{q0} and E_{q2} and also of E_{q1} and E_{q3} is a consequence of the non-equivalence of E_{r0} and E_{r2} if $r > q$. This is seen by adding $r - q$ generators of order 2^e to the q generators which give rise to E_{qm}.

Thus it remains to prove the non-equivalence of E_{r0} and E_{r2} for some arbitrarily high r. For any group H of the type under consideration, let H' be the subgroup of H determined by the elements of order 2 of H. The value of $[a, a]$ is constant on each co-set of H' in H. In the table below the number of these co-sets, on which $[a, a]$ takes specified values, is listed for five groups E_{rm}. If the matrix B, belonging to H, is E_r, the number of co-sets of H' in H, in which $[a, a] \equiv \alpha$ modulo 4, is obviously $\sum_{0 \leqq 4l+\alpha \leqq r} \binom{r}{4l+\alpha}$. This gives the first, third and fourth columns of the table. The second may be easily verified directly. The last may be obtained from the second and third by noting that the group H corresponding to $E_{r+1\,2}$ is the direct sum of the groups corresponding to E_{r0} and E_{22}. The table shows immediately that the matrices $B = E_{r0}$ and $B = E_{r2}$ are not equivalent for any r of the form $8k + 2$.

$[a, a] \diagdown B$	E_2	E_{22}	$E_{8k\,0}$	$E_{8k+2\,0}$	$E_{8k+2\,2}$
0	1	1	2^{8k-2}	$2^{8k} - 2^{4k}$	$2^{8k} + 2^{4k}$
1	2	0	$2^{8k-2} + 2^{4k-1}$	2^{8k}	2^{8k}
2	1	1	2^{8k-2}	$2^{8k} + 2^{4k}$	$2^{8k} - 2^{4k}$
3	0	2	$2^{8k-2} - 2^{4k-1}$	2^{8k}	2^{8k}

8. Let B be the matrix (11) of a group H with r generators of order $2^e > 4$, and suppose that B may be transformed into diagonal form. *Then B is equivalent to the matrix $B' = \| h_i \delta_{ij} \|$ modulo 2^e if and only if B is equivalent to B' modulo 8.* For $e = 3$ this is evident. Suppose that the statement has been proved for $e - 1$, so that B may be transformed modulo 2^{e-1} into the form B'. The same transformation transforms B modulo 2^e into the form $B' + 2^{e-1}D$, where each element of D is 0 or 1. If B is reduced to diagonal form by the method in **3**, it retains the form $B' + 2^{e-1}D$, where D is now a diagonal matrix each element of which is 0 or 1. The proof by induction of the above statement in italics is completed by the remark that $1 + 2^{e-1}$ is a quadratic residue modulo 2^e for $e > 3$, since

$$5^{2^c} \equiv 1 + 2^{c+2} \text{ modulo } 2^{c+3}, \qquad c = e - 3 \geqq 0.$$

Thus it remains to consider the diagonal matrix B if $2^e = 8$. In that case *any two diagonal elements h_1 and h_2 of B may be replaced simultaneously by $h_1 + 4$ and $h_2 + 4$.* In fact it is sufficient to replace the corresponding generators b_1 and b_2 of H by $b_1 + 2h_1b_2$ and $b_2 - 2h_2b_1$. Thus, if B is equivalent modulo 4 to the form E_{rm}, it is equivalent modulo 8 (and also modulo 2^e, $e > 3$) either to E_{rm} or to F_{rm}, where F_{rm} is obtained from E_{rm} by adding 4 to the first diagonal element. That E_{rm} and F_{rm} are not equivalent is evident, since their determinants are distinct modulo 8. Accordingly, *one has for the matrix B corresponding to a group H with r generators of order $2^e > 4$ the eight (or $2r + 2$ if $r \leqq 3$) non-equivalent diagonal normal forms E_{rm}* and *F_{rm}, $0 \leqq m \leqq 3$, $m \leqq r$.*

The following statement may be obtained easily from the above considerations: *Two matrices B and B', which belong to two groups with r generators of order $p^e = 2^e \geqq 8$ and have at least one odd diagonal element, are equivalent modulo 2^e if and only if they are equivalent modulo 8.*

9. Let B be the matrix (11) belonging to a group H with r generators b_i of order 2^e and suppose that every diagonal element of B is even, so that B cannot be transformed into diagonal form and r is even. If h_{ij} is an odd element of B, the submatrix of B formed by the elements in the intersections of the i-th and j-th column with the i-th and j-th row is non-singular modulo 2, so that the remaining elements in these rows and columns may be changed into zeros by adding multiples of b_i and b_j to the other generators of H. Continuing this process one may transform B into a diagonal matrix the elements of which are two rowed square matrices. The odd elements in these two-rowed matrices may be made equal to 1 by multiplying one of the two corresponding generators with some odd number. Then each such matrix appears in the form

$$C = \begin{pmatrix} 2k & 1 \\ 1 & 2l \end{pmatrix}. \tag{15}$$

If two matrices of the form (15) *are congruent modulo* 4, *then they are equivalent modulo* 2^e. If $e = 1$ any two matrices of the form (15) are equivalent. If $e = 2$ their equivalence modulo 2^e is an immediate consequence of their congruence modulo $4 = 2^e$. Thus the above statement will be proved by complete induction if it is shown that, if two matrices (15) are equivalent modulo 2^{e-1}, they are equivalent modulo 2^e, $e > 2$. If C is the one matrix, and the other may be transformed into C modulo 2^{e-1}, then it may be transformed into the form $C + 2^{e-1}D$ modulo 2^e, where D is symmetric and each of its elements is either 0 or 1. Now any diagonal element of D may be made equal to 0, since, if $e > 2$, one may add 2^{e-1} to a diagonal element of $C + 2^{e-1}D$ by adding to the corresponding generator 2^{e-2} times the other generator. Multiplying, if necessary, one of the generators by $1 + 2^{e-1}$ one finds that C and $C + 2^{e-1}D$ are equivalent, so that the above statement in italics follows by complete induction.

Any matrix (15) *is equivalent modulo* 4, *and hence modulo any power of* 2, *to one of the two matrices*

$$I_2 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad J_2 = \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}. \tag{16}$$

In fact, since equivalence modulo 4 is a consequence of congruence modulo 4 and generators may be interchanged, it is sufficient to show that the matrices I_2 and

$$\begin{pmatrix} 0 & 1 \\ 1 & 2 \end{pmatrix}$$

are equivalent modulo 4. This is evident if one replaces the second generator by the sum of the two generators.

The four rowed matrices

$$\begin{pmatrix} I_2 & 0 \\ 0 & I_2 \end{pmatrix} \text{ and } \begin{pmatrix} J_2 & 0 \\ 0 & J_2 \end{pmatrix}$$

are equivalent modulo 2^e. For $e = 1$ this is evident. If $e > 1$ transform the second matrix by adding the third generator of the corresponding group to the first generator. The upper left-hand corner then takes a form equivalent to I_2 and may be reduced to I_2. If the matrix is reduced further to normal form, the lower right-hand corner must assume either the form I_2 or the form J_2. That it cannot assume the form J_2 follows from the fact that I_4 and J_4 are not equivalent. This will be proved in **10**, where I_r and J_r are defined for every even r.

10. If r is even, let I_r denote the matrix obtained by arranging $\frac{1}{2}r$ matrices I_2 in the diagonal of an r-rowed matrix, of which the remaining elements are zero, and let J_r denote the matrix obtained from I_r by replacing the first of the matrices I_2 by J_2. It is easily seen that the considerations in **9** imply:

If B is the matrix (11) *of a group H with r generators of order 2^e and if every diagonal element of B is even, then r is even and B may be transformed into the form I_r or J_r.* It will next be proved that *I_r and J_r are equivalent modulo 2^e if and only if $e = 1$.* The case $e = 1$ of this statement is evident. It is also evident that the case $e > 2$ is a consequence of the case $e = 2$. It remains to show that I_r and J_r are not equivalent modulo 4. But this is evident from $\det I_r \not\equiv \det J_r$ modulo 8 and from the following statement.

If B has even diagonal elements and B is replaced by an equivalent matrix modulo $2^e = 4$, then the residual class of $\det B$ *modulo* 8 *remains unchanged.* In fact, changing B into an equivalent matrix either results in multiplying $\det B$ by the square of an odd constant (and such a square is always congruent to 1 modulo 8), or it consists in adding multiples of 4 to the elements of B. The latter changes do not affect the residual class of $\det B$ modulo 8, since B has to remain symmetric, has even diagonal elements and has an even number of rows.

The results obtained may be restated as follows: *Two matrices B and B', which belong to two groups with r generators each of order $p^e = 2^e \geqq 4$, and one of which has no odd diagonal element, are equivalent modulo 2^e if and only if they are equivalent modulo* 4.

If a matrix B belongs to a group H, which has r generators of order 2^e, $e > 1$, and if every diagonal element of B is even, then B is equivalent to I_r or J_r according as $\det B \equiv (-1)^{\frac{1}{2}r}$ *or* $\det B \equiv 5(-1)^{\frac{1}{2}r}$ *modulo* 8.

11. Let G be a group of which all elements have an order which is a power of 2 and let the corresponding matrix A (cf. **2**) be transformed into normal form by the methods of **6-10**. As in **5**, it is necessary to investigate in how far this normal form is dependent on the particular choice of the subgroups H of G corresponding to the different orders 2^e of base elements of G. By (14) and the statements at the end of **8** and **10** the invariant properties of a matrix $A^{(e)}$ derived from A, are also invariant properties of A if $e = 1$, or if $e = 2$, $c > 1$ or if $c > 2$ (Immediately below (14) the definition of c may be found).

The problem described above will be investigated here in the case where only two distinct exponents e occur in the orders 2^e of base elements of G and if the difference of these two exponents is not more than 2, i. e. if in (14), $c \leqq 2$. On the basis of the facts obtained, it will be easy to decide what are the invariant properties of any particular matrix A.

The following three statements will be needed:

(17) *Modification modulo 4 of a matrix B in the normal form I_r or J_r or modification modulo 2 of non-diagonal elements of such a matrix does not change the normal form.* The first part of this statement has already been proved. The second part depends of course on the fact that all matrices under consideration are symmetric. It is easily checked that the addition of one pair of symmetrically placed even elements to I_r or J_r does not change the residual class of the determinant modulo 8. But the same must hold if more than one such pair is added. For any new term, which contains one element each of more than one pair of new elements, cannot affect the determinant modulo 8, since it is either divisible by 4 and duplicated by a symmetric term, or else is symmetrically placed, in which case it is divisible by 16.

(18) *Modification modulo 8 of a matrix B in the normal form E_{rm} or F_{rm} or modification modulo 4 of non-diagonal elements of such a matrix does not affect the normal form.* The first part has already been proved. For the second part it should be remembered that m is invariant under modifications modulo 4. The statement follows, since it is easy to verify that the determinant modulo 8 is not changed by addition of pairs of symmetrically placed numbers which are divisible by 4.

(19) *Modification modulo 4 of a matrix B in the normal form E_{rm} or F_{rm} or modification modulo 2 of non-diagonal elements of such a matrix may interchange E and F but cannot change the value of m.* The first part is already known. For the second part, notice that if a pair of symmetrically placed even elements is added to E_{rm} or F_{rm} and this pair is eliminated by a change of base according to **3**, then the change in diagonal elements is a modification modulo 4, so that the second part is a consequence of the first part.

In what follows a matrix A is supposed to be built up out of a matrix $A^{(e)}$ and a matrix $A^{(e+1)}$ or $A^{(e+2)}$, both of which are supposed to have normal form and the problem is to determine the invariant properties of A. The large number of cases to be considered has been arranged in five groups of which only the last gives some trouble.

The form $A^{(e)}A^{(e+2)}$ *cannot be modified if at least one of the normal forms is* I_r *or* J_r. That the one equal to I_r or J_r cannot be modified follows from (14) and (17). But if one of the matrices has the form E_{rm} or F_{rm}, its normal form cannot be modified by (18), which is applicable, since the other matrix has even diagonal elements.

The form $A^{(e)}A^{(e+1)}$ *cannot be modified if both matrices have a normal form of the type* I_r *or* J_r. In fact in this case the second half of (17) is applicable, since all diagonal elements are even.

In the form $A^{(e)}A^{(e+2)}$, *where both* $A^{(e)}$ *and* $A^{(e+2)}$ *have normal forms of the type* E_{rm} *or* F_{rm}, *the two values of* m *and the product of the determinants of* $A^{(e)}$ *and* $A^{(e+2)}$ *modulo* 8 *are invariants.* In other words both $A^{(e)}$ and $A^{(e+2)}$ may change from E to F or conversely at the same time but not separately. That the two values of m are invariant follows from (14) and (19). That the product of the determinants is invariant modulo 8 is evident, since the transformation of generators for $H^{(e)}$ and $H^{(e+2)}$ combined consists in multiplication with a non-singular matrix. It remains to show that the determinant of $A^{(e+2)}$ is not invariant. This is clear, since the determinant of $A^{(e+2)}$ changes if any generator b_i of $H^{(e+2)}$ is replaced by the sum of b_i and any generator of $H^{(e)}$.

In the form $A^{(e)}A^{(e+1)}$, *where one matrix is of the type* I_r *or* J_r *and the other is of the type* E_{sm} *or* F_{sm}, *the value of* m *and the product of the two determinants modulo* 8 *are the only invariants.* In other words the one may change from I_r to J_r or conversely if the other changes at the same time from E_{sm} to F_{sm} or conversely. That m is invariant follows from (19). The product of the determinants is always invariant modulo 8. In order to see that the determinants of $A^{(e)}$ and $A^{(e+1)}$ are not separately invariant in case $A^{(e+1)}$ has the form I_r or J_r, it is sufficient to add one generator of $A^{(e)}$ to both generators corresponding to a matrix I_2 or J_2 in $A^{(e+1)}$. If $A^{(e)}$ has the type I_r or J_r, one obtains the same result by adding twice one generator of $A^{(e+1)}$ to a similar pair of generators of $A^{(e)}$.

If $e = 1$ or if $e = 2$ and $A^{(e)}$ is of type E or F in the last two statements, the product of the determinants modulo 8 is not an invariant of the matrix A, since then $\det A^{(e)}$ is not an invariant of $A^{(e)}$.

Let the two matrices $A^{(e)}$ $A^{(e+1)}$ *both have a normal form of type* E *or* F. In (20) each vertical pair of numbers represents diagonal elements the first in $A^{(e)}$, the second in $A^{(e+1)}$. Two vertical pairs separated by the sign $\sim$ are such that the one pair may be replaced by the second in the normal form under discussion. If $e = 1$ or $e = 2$ the table may be simplified in an obvious way.

$$(20)\qquad \begin{array}{lcccccccc} e: & 1\ \ 3 & 1\ \ 7 & 1\ \ 3 & 1\ \ 7 & 3\ \ 5 & 3\ \ 5 & 5\ \ 7 & 5\ \ 7 \\ & \sim & \sim & \sim & \sim & \sim & \sim & \sim & \sim \\ e+1: & 1\ \ 3 & 3\ \ 5 & 5\ \ 7 & 7\ \ 1 & 1\ \ 7 & 5\ \ 3 & 1\ \ 3 & 5\ \ 7 \end{array}$$

In order to see how (20) is obtained, consider the generators a and b of order 2^e and 2^{e+1} with $L(a, a) = k2^{-e}$, $L(a, b) = 0$, $L(b, b) = l2^{-e-1}$. The new generators $a' = a + 2b$, $b' = kb - la$ satisfy $L(a', a') = (k + 2l)2^{-e}$, $L(a', b') = 0$, $L(b', b') = (k^2l + 2l^2k)2^{-e-1}$. Since in every $A^{(e)}$ any reduction modulo 8 may be obtained by a change of generators and since k and l are odd numbers, one sees that the pair k, l may be replaced by the pair $k + 2l, l + 2k$. That no other equivalences are valid, except those in (20) may be verified easily. Now if $A^{(e)}A^{e+1}$ is a matrix A, and if both $A^{(e)}$ and $A^{(e+1)}$ are of the form E_{rm} or F_{rm}, then it is easily seen from (20) that $A^{(e+1)}$ may be reduced to the form E_{r0} or F_{r0} and also that, if $A^{(e)}$ and $A^{(e+1)}$ do not both have exactly one row, *then* $A^{(e+1)}$ *may be reduced to the form* $E_{r0} = E_r$. *The invariants of the new form of* $A^{(e)}$ *are now the invariants of the pair* $A^{(e)}\ A^{(e+1)}$. This statement may be proved by a method which is closely related to the one used at the end of **7.** Of course if both $A^{(e)}$ and $A^{(e+1)}$ have one row and $e > 1$, then the pair $A^{(e)}\ A^{(e+1)}$ has in addition the invariant property that $A^{(e+1)}$ may be reduced to the form E_{10} or F_{10}.

12. Examples may be given to show that all normal forms found above for the matrix of looping coefficients of a manifold of dimension $n = 4k + 3$ with given $2k + 1$ dimensional torsion may actually occur. The problem to prove this is simplified very much by the well known remark that the sum of two manifolds M_n and M'_n has a matrix of looping coefficients which appears in the form

$$\begin{pmatrix} L & 0 \\ 0 & L' \end{pmatrix}$$

where L and L' are the matrices of looping coefficients of M_n and M'_n. The sum of two manifolds of the same dimension n is obtained by omitting an n-simplex from both manifolds and identifying the boundaries of the remaining bounded manifolds. Thus any normal form of a matrix of looping coefficients which is diagonal may be obtained by forming the sum of suitably chosen lens spaces[8] (Linsenräume) and all other normal forms may be obtained if examples are known of manifolds with 2 torsion numbers 2^e, such that their matrices of looping coefficients have the normal forms I_2 and J_2.

In case $n = 3$ many examples which are not sums of simpler manifolds may be found among Seifert's "gefaserte Räume." Here only the manifolds of that type will be considered with a signature[9] of the form

---

[8] Compare H. Seifert, *loc. cit.*, pp. 824-825.

[9] For the definition and properties of "gefaserte Räume," see H. Seifert, *Acta*

(21) $$(O\,0; 0\,|\,0; \alpha_0, 2^e; \alpha_1, 2^e; \cdots, \alpha_r, 2^e),$$

where

(22) $$r \text{ is even, } \alpha_i \text{ is odd, } i = 0, 1, 2, \cdots, r, \ \alpha_0 = 1.$$

Manifolds with such signatures are obtained by the following construction.

Let $C \times S$ be the product of a simple closed curve C and a 2-dimensional manifold S of genus zero. Let C_i, $i = 0, \cdots, r$, be $r + 1$ simple closed curves on S of which the "interiors" have no point in common. Now omit for each i the product of C and the interiors of C_i and replace this product by a different anchor ring T_i with the same boundaries $C \times C_i$ but on which $2^e b_i + \alpha_i a$ is homologous to zero instead of b_i. Here a is a basic cycle on $C \times p$ (p is any point of S) and b_i is a basic cycle on $q \times C_i$ (q is any point of C).

Let u_i be a 2-chain which is on T_i, except for the product of S, and an arc from a point of C_i to the point p, and suppose that

(23) $$u^i \rightarrow 2^e b_i + \alpha_i a.$$

Then one has for the intersection number $S(u_i, b_j)$ of u_i and b_j,

(24) $$S(u_i, b_j) = \alpha_i \delta_{ij},$$

where b_j is thought of as moved into the interior of T_j by a small deformation. As generators of the homology group of the manifold under consideration we may use b_i and a, with the basic homologies

(25) $$2^e b_i + \alpha_i a \sim 0, \quad i = 0, \cdots, r, \quad \text{and} \quad \sum_{i=0}^{r} b_i \sim 0.$$

The last relation expresses the fact that the "exterior" of all curves C_i has as its boundary the sum of all simple closed curves C_i. As new generators of the same homology group one may use

(26) $$a \quad \text{and} \quad c_j, \quad \text{where} \quad c_j = \alpha_j b_0 - b_j, \quad j = 1, \cdots, r.$$

For (25) and (26) imply

$$\sum_{j=1}^{r} c_j \sim \beta b_0 \quad \text{and} \quad 2^e c_j \sim 0, \quad j = 1, \cdots, r, \quad \beta = \sum_{i=0}^{r} \alpha_i.$$

By (22), the coefficient β of b_0 is an odd number. Thus b_0 may be obtained as a linear combination of $\beta b_0 \sim \sum_{j=1}^{r} c_j$ and $2^e b_0 \sim -a$. Since $b_j = \alpha_j b_0 - c_j$, this implies that every element of the homology group may be obtained as a linear combination of the elements (26). The basic homology relations (25), if expressed in terms of the generators (26), take the form

(27) $$\beta a \sim 0, \quad 2^e c_j \sim 0, \quad j = 1, \cdots, r.$$

Mathematica, vol. 60 (1933), pp. 147-238. For the signature (21), see in particular p. 181. For the statement that the spaces are not sums of simpler manifolds, see p. 229.

The α_i are not required to be positive. Accordingly one may choose the α_i in such a way that $\beta = \sum_{i=0}^{r} \alpha_i = 1$, in which case the manifold has r torsion numbers equal to 2^e. Whether or not one makes this assumption, the only group H of interest has the r generators c_i of order 2^e. The corresponding matrix B may be determined as follows. By (23) and (26)

$$\alpha_i u_0 - u_i \to 2^e(\alpha_i b_0 - b_i) = 2^e c_i, \quad i = 1, \cdots, r.$$

Hence on denoting again by h_{ij}, $i, j = 1, \cdots, r$ the general element of B,

$$h_{ij} = S(\alpha_i u_0 - u_i, \alpha_j b_0 - b_j),$$

and so, by (24),

$$(28) \qquad B = \| h_{ij} \| = \| \alpha_i \alpha_j + \alpha_i \delta_{ij} \|, \quad i, j = 1, \cdots, r.$$

Since all diagonal elements of B are even by (22), the normal form of B is I_r or J_r according as

$$(29) \qquad \det B \equiv (-1)^{\frac{1}{2}r} \quad \text{or} \quad \det B \equiv 5(-1)^{\frac{1}{2}r} \text{ modulo } 8,$$

and before determining $\det B$ one may reduce the elements of B modulo 4. So suppose that

$$(30) \qquad \alpha_i \equiv 1 \text{ modulo } 4,\ i \leqq s; \quad \alpha_i \equiv -1 \text{ modulo } 4,\ i > s; \quad r - s = t.$$

Then B reduces modulo 4 to the form

$$\begin{pmatrix} A_{ij} & C_{lj} \\ C_{ik} & D_{lk} \end{pmatrix} \quad i, j = 1, \cdots, s; \quad l, k = s+1, \cdots, s+t,$$

where

$$A_{ij} = 1 + \delta_{ij}, \quad C_{lj} = C_{ik} = -1, \quad D_{lk} = 1 - \delta_{lk}.$$

In order to compute the determinant, first change the sign of the last t columns, then subtract the $(i+1)$-th column from the i-th column successively for $i = 1, \cdots, s+t-1$. Next subtract suitable multiples of the first up to the $(s+t-1)$-th column from the last column, making all elements of the last column zero except the last element. Thus one sees that

$$(31) \qquad \det B \equiv (-1)^t (s - t + 1) \text{ modulo } 8.$$

Comparing (29) and (31) the following statement may be verified:

If a manifold with signature (21) *satisfies* (22) *and* (30), *then its matrix of looping coefficients has the normal form* I_r *if* $s - t \equiv 0$ *or* 6 *modulo* 8 *and* J_r *if* $s - t \equiv 2$ *or* 4 *modulo* 8.

THE JOHNS HOPKINS UNIVERSITY.

FIXED POINTS AND THE EXTENSION OF THE HOMEOMORPHISMS OF A PLANAR GRAPH.*

By SAUNDERS MACLANE and VIRGIL W. ADKISSON.

1. Introduction. Gehman, Adkisson, and Moore[1] have found conditions under which a given homeomorphism of a continuous curve on the sphere can be extended to the sphere. In this paper we consider the problem[2] of characterizing those graphs on the sphere for which every homeomorphism of the graph can be extended to the sphere. We restrict ourselves to cyclicly connected[3] graphs and obtain a necessary and sufficient condition that such a graph has some one map[4] on the sphere in which all of its homeomorphisms are extendable. This necessary and sufficient condition depends essentially on the behavior of certain fixed points, for it asserts merely that the fixed points of all homeomorphisms of order 2 must be distributed on the graph as are the fixed points of periodic homeomorphisms of the sphere.

1. 1. Let G_1 be a cyclicly connected continuous curve lying on the sphere and consisting of a finite number of simple closed curves. Then G_1 is composed of a number of simple arcs $\alpha_1, \beta_1, \cdots$, each ending on two branch points. These simple arcs and branch points, considered as elements, constitute a *combinatorial graph* G; such a graph is made up of a finite number of *edges* (simple arcs) $\alpha, \beta, \gamma, \cdots$ and a finite number of *vertices* $p, q, r, \cdots$, each edge "ending" on just two vertices.[5] An edge α with ends p and q will be denoted by $\alpha(p, q)$. The original curve G_1 is a *map* of the graph G; for a given G_1, we may assume, without loss of generality, that each vertex of G is on at least three edges of G.

A given homeomorphism T of G_1 to itself takes each arc α_1 of G_1 into some arc $\beta_1 = T\alpha_1$ and thus sets up a one-to-one correspondence between the

---

* Presented to the Society, March 27, 1937. Received by the Editors April 8, 1937.

[1] Gehman, [1], [2], [3]; Adkisson, [1], [2]; Moore, [1]. Numbers in brackets refer to the bibliography at the end of the paper.

[2] One form of this problem was originally suggested by J. R. Kline.

[3] For this concept see Whyburn, [1], [2], [3], and Whyburn and Kuratowski [1]. A cyclicly connected graph is also non-separable in the sense of Whitney. See Whitney, [1], Theorem 7.

[4] In another paper we characterize graphs such that all homeomorphisms are extendable in every map. See Adkisson and MacLane [1].

[5] For definitions of other combinatorial terms used, see Whitney [1] or [2], or MacLane [1].

corresponding edges α and β in the combinatorial graph G. This correspondence is an automorphism of G. Here an *automorphism* σ of a combinatorial graph G is a one-to-one correspondence $\alpha \leftrightarrow \sigma\alpha$ of the edges of G to the edges of G and a one-to-one correspondence $p \leftrightarrow \sigma p$ of the vertices p of G to the vertices of G such that each edge $\alpha(p, q)$ corresponds to an edge $\sigma\alpha$ with the ends σp and σq.

The extendability of a homeomorphism is intimately related to the possibility of disconnecting the graph by removing two of its vertices. We shall call this "splitting" the graph. More explicitly, a *semi-split* of G at the vertices h_1 and h_2 is a representation of G as a sum

$$G = H + H', \tag{1}$$

where H and H' are two non-void subgraphs, having in common no arcs and no vertices except for the vertices h_1 and h_2. This semi-split is a *split* of G if neither H nor H' is a chain. A subgraph H in a semi-split (1) of G at h_1 and h_2 will be called *least* if H has itself no semi-split at h_1 and h_2; that is, if any two arcs of H are connected outside h_1 and h_2. A split (1) of G will itself be called *least* if either of the components H or H' of the split is least. If (1) is a least split, the subgraphs H_1 and H_2 are *components* of G, and each component has the *ends* h_1 and h_2. Note that component means "component of a least split." [6]

1.2. If a graph G is mapped on the sphere, and if a homeomorphism of the sphere carries G into itself, then a fixed point x of the homeomorphism on the sphere may be on the graph, in which case the graph will have a fixed vertex or a fixed edge. Otherwise x lies in one of the complementary domains into which the graph subdivides the sphere, so that the boundary of this complementary domain must go into itself in the homeomorphism. If G is cyclicly connected, this boundary is a circuit of G. Hence fixed points on the sphere appear in G either as fixed vertices, edges, or circuits. We define for any automorphism σ of G: a *fixed vertex* p is one with $\sigma p = p$. A *fixed edge* α is one with $\sigma\alpha = \alpha$, while a circuit C is a fixed circuit if $\sigma C = C$, and if C contains no fixed vertices or edges. We note that in a fixed circuit C the cyclic order of the vertices and edges must be preserved by σ, since an inversion of this order would necessarily involve two fixed points, either edges or vertices, on C. A *combinatorial fixed point* will mean either a fixed vertex, edge, or circuit.

---

[6] It can be shown that a cyclicly connected graph G has a split if and only if it contains a "split circuit," in the sense of this term used in Adkisson and MacLane, [1].

1. 3. A homeomorphism of a graph which can be extended to the sphere gives essentially[7] a periodic transformation T of the sphere into itself. But a periodic transformation is homeomorphic to a rotation about an axis or to a reflection in a diametral plane or to a product of a rotation with a reflection, and hence has either two fixed points, or a circuit all of whose points are fixed, or else no fixed points (cf. Eilenberg, [1]). Some impossible configurations would thus be

(i) a circuit of fixed points plus another fixed point;

(ii) three distinct fixed points plus a fixed circuit.[8]

Our main theorem states essentially that a graph with no such impossible configurations has a map in which all homeomorphisms are extendable. In the census of distinct fixed points for condition (ii), several fixed circuits might unfortunately correspond to the same fixed point on the sphere. In this case the fixed circuits will "nest" in some fashion. More explicitly, three fixed circuits, each pair of which lies on one side of the third, would correspond to three distinct fixed points on the sphere. In this statement, the term "side" refers to the map of the graph on the sphere. We can replace it by a purely combinatorial term. A *combinatorial side* of a circuit C in G is a subgraph $H \subset G$ such that G has a least split $G = H + (G - H)$ in which[9] $C \subset (G - H)$.

We thus obtain the following condition which refers only to combinatorial properties of the graph. In stating this condition an exception is made of the particular graph G_0, consisting of two vertices p and q, and of four distinct arcs, each with the ends p and q.

THEOREM 1. *If G is a cyclicly connected planar graph, not the graph G_0 defined above, and if each vertex of G is on at least three edges, then G can be so mapped on the sphere that any homeomorphism of G into itself can be extended to a homeomorphism of the sphere into itself if and only if there is no automorphism σ of G with $\sigma^2 = 1$ and with either one of the following invariant configurations in G:*

---

[7] More explicitly, T^n for some n will carry every complementary domain of the graph into itself.

[8] The presence of a fixed circuit excludes the possibility of representing σ by a reflection of the sphere. For, if σ is of order 2, the assertion that there is a fixed circuit is equivalent to the assertion that there is a chain A containing neither fixed vertices nor fixed edges and having the ends p and σp. This means that there is some vertex p not separated from its image vertex σp by the set S of all fixed edges and vertices, while in a reflection in a diametral plane the fixed edges and vertices do separate every vertex in one hemisphere from its image in the other hemisphere.

[9] $G - H$ denotes the subgraph of G containing all edges of G not in H.

(i) *A circuit C, all of whose vertices and edges are invariant, and a fixed edge or vertex not on C;*

(ii) *A fixed circuit C and three distinct combinatorial fixed points, x, y, and z, such that if any one of x, y, or z is a circuit*[10] *D, then the other two of x, y, and z are both contained in one combinatorial side of this circuit D.*

1. 4. Since the condition of Theorem 1 refers only to automorphisms of order 2, it will have the following curious consequence.

THEOREM 2. *If a planar combinatorial graph G is cyclicly connected and has each vertex on at least three edges, and if G has no automorphisms of order 2, then G has a map in which every homeomorphism is extendable to the sphere.*

It is more remarkable that the condition of Theorem 1 refers only to properties of individual homeomorphisms and not to the relation between several homeomorphisms. This can be stated as follows:

THEOREM 3. *If the planar combinatorial graph G is cyclicly connected, has each edge on three or more vertices, and if G has no map on the sphere in which every homeomorphism is extendable, then there is one homeomorphism T of G which can not be extended in any map of G.*

This theorem is true even for the graph G_0, for here the automorphism cyclicly permuting three edges and leaving the fourth edge fixed is not extendable in any map.[11]

1. 5. The proof of Theorem 1 will first be reduced to a purely combinatorial problem. This will be done in § 2, where we also treat a simpler but analogous problem which shows the essential rôle of the "splits." The treatment of the combinatorial problem will then proceed by mathematical induction. The essential device consists in obtaining from a given graph G a simpler graph, the "skeleton" of G, which has all of the homeomorphisms of G. The existence and properties of this skeleton will be discussed in §§ 4 and 5, while the subsequent paragraphs will show by a suitable subdivision into cases that the map of the skeleton can generate a corresponding map of

---

[10] If x is a circuit, it is a fixed circuit, so that the presence of the fixed circuit C need not be explicitly required.

[11] For any $G \neq G_0$, the homeomorphism T of Theorem 3 can be taken to be of order 2, but for G_0 there is no such "nowhere extendable" automorphism of order 2. This unique behavior of the graph G_0 justifies the apparently capricious exceptional rôle of this graph in our main theorem.

the original graph in which the homeomorphisms will be extendable. This proof of the sufficiency of the condition of Theorem 1 is supplemented in § 3 by a simple topological proof for the necessity of the condition.

2. Splits versus extendability. In this paragraph we consider the problem of the extendability of a homeomorphism which carries a graph into some other graph, with the following conclusion:

THEOREM 4. *If G_1 and G_2 are two maps of a cyclicly connected combinatorial graph G on the spheres S_1 and S_2 respectively, then every homeomorphism T of G_1 to G_2 can be extended to a homeomorphism of the sphere S_1 to S_2 if and only if G has no split.*

2.1. A given homeomorphism T establishes an automorphism of G, provided each vertex of G is on at least three edges. For if α_1 is any arc of G_1, and if α_1 and $T\alpha_1$ are maps of the edges α' and α'' respectively of G, then the correspondence $\sigma\alpha' = \alpha''$ yields an automorphism of G.

The extendability of a homeomorphism T to the sphere depends upon combinatorial properties of the corresponding automorphism. The boundaries of the complementary domains into which the cyclicly connected graph G_1 divides the sphere (call them *c. d. boundaries*) are the maps of circuits of G.[12] These circuits control extendability as follows:

THEOREM 5. *A homeomorphism T of a cyclicly connected graph G_1 on a sphere S_1 to a graph G_2 on a sphere S_2 can be extended to a homeomorphism of S_1 to S_2 if and only if T carries every c. d. boundary of G_1 to a c. d. boundary of G_2.*

This theorem is an easy extension of Theorem 2 of Adkisson, [1], which was proven from a theorem of Gehman, [1]. In this case, since the number of domains is finite, we can also proceed thus: Let R_1 and R_2 be regions of S_1 and S_2 respectively whose boundaries C_1 and C_2 correspond under σ_1. Use the result that the interior and boundary of a Jordan curve can be mapped on the interior and boundary of a circle to extend σ_1 so that it carries R_1 into R_2. The combination of these extensions is a homeomorphism of S_1 to S_2.

2.2. If a graph G has no splits the c. d. boundaries of its maps on the sphere are uniquely determined. For these boundaries are simply[13] the circuits C in G whose removal from G does not disconnect G. Hence any automorphism must carry these circuits into themselves, so that any homeomorphism of G_1 necessarily preserves the c. d. boundaries and is therefore extendable by Theorem 5. This proves the sufficiency of the condition of Theorem 4.

---

[12] See Whyburn, [1], Theorem 10 or MacLane, [2], Theorem 5.3. The latter theorem gives a complete combinatorial characterization of these circuits.

[13] G is either triply connected or a θ-graph. For a θ-graph, the statement above about c. d. boundaries is immediate. For a triply connected graph it is given in MacLane, [3], Theorem 6.

2.3. For the necessity proof we need the fact that in a map of a split graph any component can be "turned around" on the two split vertices. If G is split at p and q into components H_1 and H_2, we introduce two new arcs $\alpha_1(p, q)$ and $\alpha_2(p, q)$, and form $H_1 + \alpha_1$, $H_2 + \alpha_2$. In any map of $H_1 + \alpha_1$ there must be two c. d. boundaries containing α_1, so that these boundaries may be written in the form $L_1 + \alpha_1$ and $M_1 + \alpha_1$, where L_1 and M_1 are two chains in H_1 with ends p and q.

LEMMA 2.4. *If $G = H_1 + H_2$ is a split at p and q, and if $H_1 + \alpha_1(p, q)$ has a map with the c. d. boundaries*

$$C_1, C_2, \cdots, C_s, \quad L_1 + \alpha_1, M_1 + \alpha_1, \quad (L_1 \subset H_1, M_1 \subset H_1) \tag{2}$$

while $H_2 + \alpha_2(p, q)$ has a map on the sphere with the c. d. boundaries

$$D_1, D_2, \cdots, D_t, \quad L_2 + \alpha_2, M_2 + \alpha_2, \quad (L_2 \subset H_2, M_2 \subset H_2), \tag{3}$$

then there is a map of G in which the c. d. boundaries are the circuits

$$C_1, C_2, \cdots, C_s, \quad D_1, D_2, \cdots, D_t, \quad L_1 + L_2, M_1 + M_2, \tag{4}$$

and another map of G with the c. d. boundaries

$$C_1, C_2, \cdots, C_s, \quad D_1, D_2, \cdots, D_t, \quad L_1 + M_2, M_1 + L_2. \tag{5}$$

Proof. In the map of H_2 on the sphere, remove α_2 and insert in its place[14] the given map of H_1. This gives one of the maps (4) and (5). The same process, with a change in notation, gives the second map.

The two maps (4) and (5) are distinct, for were these sets of c. d. boundaries identical, we should have either $M_1 = L_1$ or $M_2 = L_2$. But $M_1 = L_1$ means that the chain $M_1 = L_1$ is the whole of H_1, contrary to the definition of a split. Therefore, the homeomorphism of G which carries the first map of any edge into the second map of the same edge does not preserve c. d. boundaries, and so is not extendable. Hence all homeomorphisms are extendable only if G has no split, as asserted in Theorem 4.

---

[14] For details, cf. MacLane, [3], Lemma 5.

3. The necessity proof. The proof that the combinatorial conditions of our main theorem are necessary requires first a demonstration of the relevance of the term "combinatorial side" used in our theorem. We shall show that a combinatorial side always lies within a topological side.

3.1. LEMMA. *If H is a combinatorial side of a circuit C in G, then in any map of G on the sphere all of H, except perhaps for two of its vertices, lies on one side of the map of C.*

Proof. By definition of a combinatorial side H, G has a least split $G = H + (G - H)$ at two vertices h_1 and h_2, while $C \subset (G - H)$. Therefore C can meet H at most in the ends h_1 and h_2 of H. Suppose that H is not all on one side of C.

In the given split, either H or $G - H$ is least. If H is least, and if E_1 is the part of H on one side of C, then $H = E_1 + (H - E_1)$ is a semi-split of H_1 because these two non-void parts can meet only in the vertices h_1 and h_2. This contradicts the assumption that H is least.

Suppose then that $G - H$ is least. If α and β are edges of H on opposite sides of C, then, because G is cyclicly connected, there is a circuit D in G containing α and β. This D must cross C to get from α to β. But α and β are in H, so that D, which has vertices on C and hence in $G - H$, must pass through the split vertices h_1 and h_2. Therefore D crosses C at h_1 and h_2 and lies entirely in H. Because $G - H$ is least, $G - H$ must as in the previous case lie entirely on one side of D. Hence C, which is in $G - H$, also lies on one side of D, contrary to the fact that D crosses C. This contradiction shows that H must lie on one side of C and can meet C at most in its end vertices h_1 and h_2. The lemma is established.

3.2. We now show that the conditions (i) and (ii) of Theorem 1 are necessary if the graph G has a map in which every homeomorphism is extendable. For suppose that G has such a map in which each edge α has a map $M\alpha$ and also that G has an automorphism σ of order 2. For this σ one can readily construct a corresponding homeomorphism T which will also have period 2 and which will have the same fixed points as σ. More explicitly, each fixed vertex of σ is fixed under T, each fixed edge α of σ is point-wise fixed or contains a fixed point under T according as σ leaves fixed or interchanges the ends of α, and each fixed circuit C maps into a simple closed curve MC such that $T(MC) = MC$ and such that T has no fixed points on MC. By assumption the transformation T can be extended to the sphere. Therefore, as in Theorem 5, T carries every c. d. boundary of the map M into a c. d. boundary of this map, while T certainly has period 2 when considered as a permutation of these boundaries. T can then be readily extended to a homeomorphism T^* of the whole sphere in such a way that T^* has the same period 2.

Suppose now that σ has the impossible configuration (i) of Theorem 1: A circuit C of edges and vertices all of which are fixed and a fixed edge or vertex not on C. Corresponding to these, T will have a simple closed curve $M(C)$ consisting of fixed points and also a fixed point not on C. But this is impossible because T^* is a periodic homeomorphism of the sphere, for the Fundamental Lemma of Eilenberg, [1], states that a periodic transformation of the surface of the sphere has either (1°) no fixed points, (2°) all fixed points, (3°) a simple closed curve of fixed points and no other fixed points, (4°) exactly two fixed points.

In the second place consider an automorph σ with the fixed circuit C and the combinatorial fixed points x, y, and z of condition (ii) of the theorem. If x, y, and z are all fixed edges or vertices, then T will have three distinct fixed points. By the result of Eilenberg quoted above this is possible only if T^* has a simple closed curve of fixed points. In this case (Eilenberg, [1], Theorem 1), T^* is homeomorphic to a reflection of the sphere in a diametral plane. But the map of the given fixed circuit C cannot cross this diametral plane, for C contains neither fixed edges nor vertices. Thus $M(C)$ lies on one side of the diametral plane, contrary to the fact that $\sigma C = C$.

In the remaining cases one or more of the given fixed points x, y, and z may be fixed circuits. If x is a circuit D, then y and z lie on a combinatorial side of D. Lemma 3.1 then shows that $M(y)$ and $M(z)$ lie on one side of $M(D)$. Call the side of $M(D)$ not containing $M(y)$ and $M(z)$ the "inside" of $M(D)$. Since σ leaves fixed y and z "outside" D, the extended T^* must carry the inside of $M(D)$ into itself. But a periodic transformation of a circle and its interior is homeomorphic either to a rotation or to a symmetry in a diameter (Eilenberg, [1], Theorem 2). Because D has neither fixed edges nor vertices, T^* is not a symmetry in a diameter of D, and so must be homeomorphic to a rotation of the inside of $M(D)$, so that T^* has a fixed point "inside" $x = D$. This fixed point is distinct from y and z. If y or z is a circuit, T^* likewise has a fixed point within this circuit, so that in any event T^* has three distinct fixed points. This gives a contradiction, as in the previous case, and this contradiction establishes the necessity of the conditions of Theorem 1.

This necessity could have been established without appeal to the general theorem of Eilenberg, the proof of which involves complexities of a point-set-theoretic character. In the present case the given periodic transformation T^* of the sphere is associated with a "triangulation" G, such that T^* is combinatorially periodic on this triangulation. This fact can be used to give for our theorem a proof depending on a readily established combinatorial substitute [15] for Eilenberg's theorem.

4. The construction of the skeleton. The sufficiency proof of our theorem proceeds by an induction, reducing a given graph to a simpler graph having essentially the same group of automorphisms. This simpler graph, called a "skeleton," will be obtained by replacing a "minimal" component of the original graph by a single edge and by simultaneously replacing all homologous minimal components by new edges.

4.1. A component M of a least split of G is *minimal* if and only if in every least split $G = H + H'$ of G the component M is contained entirely in H or entirely in H'. We now deduce several requisite properties of minimal components. The definition immediately proves the property

4.2. *Two minimal components are either identical or else have at most end vertices in common.*

4.21. *In any least split $G = E_1 + E_2$ at the vertices e_1 and e_2 the component E_1 is either cyclicly connected or else E_1 contains another component of G with fewer edges than E_1.* For, if E_1 is not cyclicly connected, then E_1 has a "cut vertex" s such that there is a "separation" $E_1 = F_1 + F_2$, in which the non-void subgraphs F_1 and F_2 have only the cut vertex s in common. We shall show that one of the F's is a component of G. F_1 can touch the rest of G at most in the vertices s, e_1, and e_2. If F_1 contains only one of these three vertices, it is cut from $G - F_1$ by this vertex, contrary to the cyclic connectivity of G. If F_1 contains all three of these vertices, only one of them, s, can be in F_2, and this is similarly impossible. Therefore s, e_1, and e_2, are distinct, and F_1 touches the rest of G at just two vertices s and e_1. Hence $G = E_2 + F_1 + F_2$, where the only common vertices of these subgraphs are e_1 and e_2 in E_2, s and e_1 in F_1, s and e_2 in F_2. This gives two semi-splits

---

[15] This substitute runs as follows: Let G be a cyclicly connected graph mapped on the sphere such that (i) each vertex of G is on at least three edges; (ii) σ is a homeomorphism of the sphere to itself with $\sigma G = G$; (iii) σ^2 but not σ carries every edge and vertex of G into itself; (iv) C is a c. d. boundary of the map of G, $\sigma C = C$, and σ preserves cyclic order on C. Then there is in addition to C either an edge or a vertex or a c. d. boundary of G which is carried into itself by σ, and there is only one such invariant element (vertex, edge, or boundary). Furthermore, if D is a c. d. boundary with $\sigma D = D$, then σ preserves cyclic order on D.

$$G = (E_2 + F_1) + F_2, \qquad G = (E_2 + F_2) + F_1. \tag{1}$$

If neither were a split, both F_1 and F_2 would be chains, so that E_1 would be a chain, contrary to assumption. Both $E_2 + F_1$ and $E_2 + F_2$ are least because each has a cut vertex (at e_1 or e_2 respectively). Therefore one of the decompositions (1) is a least split, so that either F_1 or F_2 is a component of G contained in E_1, as asserted.

4. 22. *Every minimal component of a cyclicly connected graph is cyclicly connected.* For a component not cyclicly connected contains by the above argument a proper subcomponent of G and hence cannot be minimal according to our definition.

4. 23. *Every component E_1 of a least split $G = E_1 + E_2$ contains a minimal component of G.* We shall show that E_1 is either minimal or contains a component of G with fewer edges than E_1. From this the assertion will follow by induction. We denote the ends of E_1 by e_1 and e_2. By § 4. 21, we may assume that E_1 is cyclicly connected. Furthermore, we may assume E_1 least, for if E_1 is not least, then either E_1 contains a smaller least component with the same ends e_1 and e_2, as required for our induction, or else E_1 consists of $t \geqq 2$ suspended chains with ends e_1 and e_2, while E_2 is least. In this case, the given split $G = E_1 + E_2$ is the only least split of G at e_1 and e_2, while in any other least split of G at vertices p and q, neither p nor q could lie inside any suspended chain of E_1, so that p and q do not disconnect E_1. In any least split of G, E_1 is, therefore, contained in one component, so that E_1 is in fact minimal.

In the remaining cases we have an E_1 which is least and cyclicly connected. If E_1 is not minimal, there is a split $G = H_1 + H_2$ at h_1 and h_2 such that E_1 is in neither H_1 nor H_2. Since arcs of E_1 are disconnected by h_1 and h_2, and since E_1 is cyclicly connected, both of h_1 and h_2 must be in E_1. As E_1 is least, one of h_1 and h_2 is not an end of E_1. Either E_2 contains neither h_1 nor h_2, or else E_2 contains one of h_1 or h_2 at one end of E_2. But all edges of E_2 are connected to both ends e_1 and e_2, so that E_2 will in no event be disconnected by h_1 and h_2. Therefore E_2 is all contained in one component of the split $G = H_1 + H_2$, say in H_1. Hence $H_2 \subset E_1$. Because H_2 and E_1 have distinct ends, H_2 has fewer edges than E_1, so that we have the component H_2 contained in E_1, as required.

4. 24. Let E be a component from a least split of G at e_1 and e_2. Form the graph $E + \alpha$ where $\alpha(e_1, e_2)$ is a new edge not in G. *Then E is minimal if and only if* (i) *each end of E is on at least two edges of E*; (ii) *$E + \alpha$ has*

no least split. This fact will show that the decision whether a given component E is minimal or not depends only on the structure of $E + \alpha$ and not on the rest of G. This graph $E + \alpha$ can be considered as a "homomorphic" image of G obtained by replacing all of $G - E$ by a single edge α with the same ends. It is then geometrically apparent and it can be readily verified combinatorially that any least split of $E + \alpha$ will yield a corresponding least split of G, if we replace α by $G - E$. Conversely, any least split of G at two vertices p and q will yield a "homomorphic" least semi-split of $E + \alpha$, provided both p and q are in E, and provided E is not all contained in one component of the given least split.

By this means we prove the necessity of the conditions above: (ii) is requisite, because any least split of $E + \alpha$ would give a homomorphic split of G in which E did not all lie in one component, so that E would not be minimal. (i) is necessary because E, to be minimal, must by § 4. 22 be cyclicly connected. Conversely, suppose (i) and (ii) to hold. Then E must be cyclicly connected, for were E to have a separation at a cut vertex s, then $E + \alpha$ must have a corresponding least semi-split which would be a split because of (i). This split would contradict (ii). Next, if E were not minimal there would be a least split $G = H + H'$ at vertices p and q, such that both H and H' contain edges of E. Since E is cyclicly connected, both the vertices p and q of this split must be in E. This split therefore gives a homomorphic least split of $E + \alpha$ at p and q, contrary to condition (ii). Therefore conditions (i) and (ii) are sufficient to make E minimal.

4. 3. Two types of minimal components may be found by classifying graphs $M + \alpha$ with no least splits as "branch graphs" and "triply connected graphs." A graph consisting only of $k \geqq 2$ suspended chains, $A_1, A_2, \cdots, A_k$ having the same two ends p and q, but not having any other vertices in common, will be called a *branch graph.* A cyclicly connected graph G may be called *triply connected* if it is not a branch graph and if it has no split. The definitions readily show that a graph with a split at two vertices must have a least split at the same vertices unless both original components were branch graphs. We deduce thence the following dichotomy:

4. 31. *A cyclicly connected graph which has no least split is either triply connected or is a branch graph.* This fact, combined with § 4. 24, then yields

4. 32. *If M is a least and minimal component, then $M + \alpha$ is triply connected, while if the minimal component M is not least, then M is a branch graph with ends m_1 and m_2, and $G - M$ is least.*

4.4. Consider now a cyclicly connected graph G with a group Σ of automorphisms and with some least split. Because of the existence of minimal components (§ 4.23), we can then find a minimal component $M = M_0$ in G. Any automorphism of the group must carry M_0 into another subgraph having the same combinatorial properties, that is, into another minimal component of G. Let

$$(2) \qquad M_0, M_1, M_2, \cdots, M_{k-1}, \qquad (M_i = \sigma M_0, \sigma \epsilon \Sigma)$$

be all distinct minimal components which arise in this fashion from M_0 by the automorphisms of Σ. These components, by § 4.2, have at most end vertices in common. Hence we can construct a new graph G' by replacing each M_i of (2) by a new arc μ_i with the same ends as M_i. The resulting graph G' we call a *skeleton* of G.

Since least splits of G never take place inside the minimal components M_i replaced in the formation of G', it is clear that least splits of G yield corresponding semi-splits of G', and conversely, as follows:

4.41. *If $G = H_1 + H_2$ is a least split of G at p and q, then G' has at p and q a least semi-split $G' = H'_1 + H'_2$, where each H'_j arises from H_j $(j = 1, 2)$ by replacing each M_i in H_j by the corresponding new arc μ_i.*

4.42. *If $G' = F'_1 + F'_2$ is a split of the skeleton G' at p and q, then G has at p and q a split $G = F_1 + F_2$, where each subgraph F_j is obtained from F'_j $(j = 1, 2)$ by replacing each new arc μ_i in F'_j by the corresponding M_i. If the given split of G' is least, then the resulting split of G is least.*

4.43. *A skeleton of a cyclicly connected graph is cyclicly connected.* This fact is readily established from the definitions.

4.5. Corresponding to the given group Σ of automorphs of G, we now define a group Σ' of automorphs of the skeleton G'. Each $\sigma \epsilon \Sigma$ induces a corresponding transformation σ' of G'. For, each edge μ_i of G' corresponds to a component M_i of G, and σM_i must be some one of the set (2) of homologous minimal components. If $\sigma M_i = M_j$, we set $\sigma' \mu_i = \mu_j$ for the corresponding new edges μ_i and μ_j, while if α is an edge of G' which is also in G we set $\sigma' \alpha = \sigma \alpha$. These equations, with $\sigma' p = \sigma p$ for any vertex p of G', define the induced transformation σ' on G'. This σ' is a $1 - 1$ transformation carrying vertices and edges of G' into vertices and edges of G'. It is an automorphism, because σ' is so constructed that an edge $\alpha(p, q)$ has an image $\sigma' \alpha$ with ends $\sigma' p$ and $\sigma' q$. Finally, the correspondence $\sigma \rightarrow \sigma'$ preserves the product of automorphs, for if $\rho = \sigma \tau$ holds in Σ, the definition shows that $\rho' = \sigma' \tau'$. Since the product is preserved the set Σ' of all these σ' is a group and the correspondence $\sigma \rightarrow \sigma'$ is a homomorphism of Σ to Σ'. We shall subsequently show that this homomorphism is in certain cases an isomorphism.

4. 6. The skeleton which we have thus defined can be used in particular cases to explicitly construct all homeomorphisms of a graph, for we need only find the group of all homeomorphisms of the skeleton (which may be constructed by replacing simultaneously all minimal components) and investigate which of these can, in fact, be extended to be homeomorphisms of the graph. By taking repeated skeletons, the problem is thus reduced to that of finding homeomorphisms of a triply connected graph.

5. Admissible groups of automorphisms. The sufficiency proof for our main Theorem 1 proceeds by induction, using the skeleton, and requires conditions on the automorphisms of the skeleton which are in stronger form than the conditions in the theorem itself. A group of automorphisms satisfying these conditions (cf. I and II below) will be called "admissible," and we shall prove in the subsequent paragraphs that any G can be so mapped that all the automorphs of a given admissible group of automorphs can be extended to the sphere.

5. 1. To define an admissible group, we consider automorphisms which "twist" a single component or which permute several components "in parallel." An automorphism σ is said to *twist* G if there is a least split of G with a component H such that $\sigma H = H$ and σ carries each end of H into itself, while σ is not the identity on H.

A representation of G with more than two subgraphs

$$G = H_1 + \cdots + H_m \qquad (H_i \neq 0, i = 1, \cdots, m) \tag{1}$$

is called a *multiple split* of G at the points p and q if any two distinct subgraphs H_i and H_j have in common no edges and only the vertices p and q, and if no H_i has a semi-split at p and q. An automorphism σ is a *parallel permutation* if there is a multiple split (1) at p and q such that $\sigma p = p$, $\sigma q = q$, $\sigma H_1 = H_2$ and such that either $m \geqq 4$ or else $m = 3$ and H_3 is not a chain.

A group Σ of automorphisms of a cyclicly connected graph G is *admissible* if and only if

(I) Σ contains no parallel permutation and no twist of order 2.

(II) Σ contains no σ of order 2 for which G contains a fixed circuit and three distinct combinatorial fixed points, x, y, and z, where x is a circuit only if y and z are in one combinatorial side of x, with similar conditions when y or z is a circuit.

5. 2. LEMMA. *If G satisfies the conditions of Theorem 1, then the group Σ of all automorphisms of G is admissible.*

The proof uses primarily the fact that Σ contains all the automorphs. The condition (ii) of the Theorem is the same as (II) of the definition. Hence we need only consider the condition (I). Suppose first, contrary to (I), that Σ were to contain a σ of order 2, twisting a component H. The transformation τ which is equal to σ on H and which is the identity on $G - H$ must then be an automorph of G. This τ is of order 2. Since $G - H$ is not a chain it contains a circuit whose vertices and edges are all invariant under τ. If $G - H$ contains edges not in this circuit these edges are also fixed under τ and τ violates the condition (i) of Theorem 1. Therefore $G - H$ must be just a circuit, and this circuit must pass through both ends p and q of H, so that $G - \mathrm{H}$ consists merely of two suspended chains L and N with ends p and q. Since each vertex of G is on at least three edges, L is a single arc, as is N. Define a new automorphism ρ which is the identity on H and which interchanges L and N. As before, H contains a circuit of points fixed under ρ. Hence, by the argument used for τ, H must consist only of this circuit. $G = H + L + N$ consists of four arcs with the same end points p and q, so that G is the particular graph G_0 excluded in Theorem 1. This contradiction shows that G can have no twists of order 2.

Suppose on the other hand that G has a parallel permutation σ, defined as above. Introduce a new automorphism τ which is equal to σ on H_1 and to σ^{-1} on H_2, and which is the identity elsewhere. This τ is of order 2 and may be treated exactly as the τ in the preceding paragraph. This treatment establishes the lemma.

5. 3. LEMMA. *If Σ is an admissible group of automorphisms for the cyclicly connected graph G and if G' is a skeleton of G with a corresponding group Σ' of automorphisms, then the correspondence $\sigma \rightarrow \sigma'$ defined in* § 4. 5 *makes Σ isomorphic to Σ'.*

The correspondence will be an isomorphism if no $\sigma \neq 1$ corresponds to the identical automorphism of Σ'. Suppose contrarywise that $\sigma \neq 1$ corresponds to $\sigma' = 1$. This means that $\sigma M_i = M_i$ for each minimal component M_i used in § 4, (2), to construct the skeleton, while σ is the identity except on these M_i. Consider the two types of minimal components established in § 4. 32.

In case M_i is a branch graph, then each one of the M's must be a branch graph. Select some one, say M_1, on which σ is not the identity and let M_1 consist of $k \geqq 2$ suspended chains $A_1, \cdots, A_k$ with ends p and q. Then σ must permute at least two of these A's, so that σ is a parallel permutation, contrary to the definition of an admissible group.

There remains the case when each one of the components M_i is least. Consider one such component M_1 with ends p and q, and adjoin a new arc $\alpha(p, q)$. Then by § 4. 32, $M_1 + \alpha$ is triply connected, so that all maps of $M_1 + \alpha$ have the same c. d. boundaries by § 2. 2. There will be two such boundaries $A + \alpha$ and $B + \alpha$ containing α. If we extend σ by setting $\sigma\alpha = \alpha$, then σ is an automorphism of $M_1 + \alpha$ which carries the c. d. boundaries into c. d. boundaries and therefore takes the unordered pair $(A + \alpha, B + \alpha)$ into itself. One of the following alternatives holds:

$$\sigma A = A, \ \sigma B = B; \quad \text{or} \quad \sigma A = B, \ \sigma B = A.$$

Under the first alternative, σ must leave every edge and vertex of these circuits $A + \alpha$ and $B + \alpha$ invariant. But σ can be extended from $M_1 + \alpha$ to a periodic homeomorphism of a sphere into itself. Because σ has more than a circuit of fixed points, it must by Eilenberg's results be the identity on $M_1 + \alpha$. In the second alternative, $\sigma^2 A = A$, $\sigma^2 B = B$, so σ^2 is, as in the previous alternative, the identity on M_1. σ^2 is thus the identity on every M_i, and so σ must actually be of order 2. But we have just shown that σ twists at least one M_i, which contradicts the condition (I) of an admissible group. Hence $\sigma \neq 1$ implies $\sigma' \neq 1$, and Σ' is isomorphic to Σ.

5. 4. LEMMA. *If Σ is an admissible group for the cyclicly connected graph G, and if G' is a skeleton of G, then the corresponding group Σ' is an admissible group on the skeleton G'.*

To show that Σ' satisfies condition (I), suppose first that Σ' contains a σ' of order 2 which twists a component H'. Then the corresponding σ is by the preceding lemma also of order 2, while the component H of the original G corresponding to H' as in § 4. 42 is itself twisted by σ, contrary to the assumption that the group Σ is admissible.

Secondly, suppose that Σ' contains a parallel permutation σ' for the multiple split

$$G' = H'_1 + \cdots + H'_m.$$

Then the argument of § 4. 42 yields a corresponding decomposition

$$G = H_1 + \cdots + H_m$$

which is a multiple split of the original G. Since $\sigma' H'_1 = H'_2$, it follows that $\sigma H_1 = H_2$ for the corresponding automorph σ, so that σ is a parallel permutation, contrary to the assumption on Σ. This contradiction shows that the group Σ' satisfies the condition (I) for an admissible group. For the condition (II) we must consider how the ends of a minimal component can be interchanged.

5.41. LEMMA. *If M is a minimal component of G with ends m_1 and m_2, and if σ is an automorphism of order 2 with $\sigma M = M$, $\sigma m_1 = m_2$, then M contains either a fixed circuit under σ or else two fixed points under σ which are either vertices or edges.*

Consider the two possible types for M indicated in § 4.32. In case M is a branch graph, let A_1 denote one of the suspended chains of M. Then either $\sigma A_1 = A_1$, in which case A_1 contains either a fixed edge or vertex, or else $\sigma A_1 \neq A_1$, in which case $A_1 + \sigma A_1$ is itself a fixed circuit.

In case M is least then, as in the proof of the Lemma 5.3, any map of $M + \alpha$ contains two c. d. boundaries $A + \alpha$ and $B + \alpha$ which can at most be interchanged by σ. If $\sigma A = A$, then the hypothesis that the ends m_1 and m_2 of A are interchanged implies that A contains a fixed edge or vertex, while $\sigma B = B$, so that B contains a second fixed point. Otherwise $\sigma A = B$, and $A + B$ is a circuit[16] which is a fixed circuit, so that in all cases M contains a fixed point.

5.42. Suppose now that condition (II) fails for the group Σ'. Then Σ' contains an automorph σ' of order 2 with a fixed circuit C' and three combinatorial fixed points x', y', and z', restricted as in (II). We will obtain a contradiction by constructing corresponding fixed points in G.

In the given C' let p be any vertex and $\sigma' p$ its image, and let A' be a chain of C' with ends p and $\sigma' p$. Replace every "new" edge μ_i in A' by any chain N_i with the same ends as μ_i and taken from the corresponding minimal component M_i. These replacements in A' yield a chain A in G with ends p and σp. $A + \sigma A$ is readily seen to be a circuit C which is fixed under σ.

If the given fixed point x' is a circuit, the same treatment will yield a corresponding fixed circuit x in G. If x' is a vertex, the same vertex $x = x'$ is fixed in G. If x' is an edge α, then either α is also an edge of G and so a fixed point x of G, or else α is one of the new edges μ_i representing a component M_i of G. Since $\sigma'\alpha = \alpha$, we then have $\sigma M_i = M_i$. If σ leaves fixed each end of M_i, these ends are fixed points in G. Otherwise, σ interchanges the ends, so that by § 5. 41 M_i does contain a fixed point x.

---

[16] For if $A + B$ is not a circuit A and B have in common a vertex s, $s \neq m_1$, $s \neq m_2$. An inspection of the map of $M + \alpha$, in which $A + \alpha$ and $B + \alpha$ are c. d. boundaries, shows that s must be a cut vertex of M, contrary to the cyclic connectivity of M (see § 4.22).

The given fixed points x', y', and z' of G' thus yield three distinct fixed points x, y, and z under σ in G. Furthermore, if x is a circuit, then y and z are on a combinatorial side of x. For the construction above gives a fixed circuit x only when x' is a circuit or when x' is an edge μ_i. If x' is a circuit, then, by the assumption that (II) is violated in G', y' and z' lie on a combinatorial side H' of x'. If all the edges μ_i in this component H' are replaced by the corresponding subgraphs M_i, we get as in § 4. 42 a corresponding "side" of x which contains the constructed fixed points y and z, as desired. In the other case, when x' is an edge μ_i, the new fixed circuit x lies in the corresponding minimal component M_i, and $G - M_i$ is a combinatorial side of x containing y and z.

For the automorph σ of order 2 in Σ we thus have a fixed circuit C and three distinct fixed points x, y, and z, contradicting condition (II) for Σ. This demonstrates the admissibility of Σ' asserted in Lemma 5. 4.

5. 5. To show the extent of the conditions for admissibility we note that they imply that an admissible group Σ contains no twist of any order, although the definition itself explicitly excludes only twists of order 2. In fact, the isomorphism of the groups Σ' and Σ shows that a twist of a given order must remain a twist of the same order on the skeleton,[17] so that any twist must eventually reduce to a twist of a minimal component in some skeleton, while such twists are either of order 2 or are parallel permutations and so are excluded.

6. Ambiguous automorphs. Certain automorphs θ in an admissible group may have so few fixed points that they can be realized geometrically either as rotations or as reflections. Such automorphs will be called "ambiguous."

6. 1. *Definition.* An automorph θ of order 2 in an admissible group Σ of G is ambiguous on the edge α of G if $\theta(\alpha) = \alpha$, if θ has no fixed circuits in G and if θ has in G exactly two combinatorial fixed points.

The fixed points of a given ambiguous θ will now be shown to divide G into two equal "halves," E_p and E_q. First, both ends of α can not be fixed points, for that would yield three fixed points. Hence θ turns $\alpha = \alpha(p, q)$

---

[17] Provided we use a skeleton with respect to the subgroup of Σ composed of the powers of the given twist.

end for end and has exactly one fixed edge or vertex x in addition to α. Let E_p be the subgraph containing all those edges β of G for which there exists a chain A_β free of fixed points, containing β, and ending on p. Define E_q similarly. Here a chain A_β is "free of fixed points" if and only if A_β contains no fixed edges and no fixed vertices, except perhaps for a fixed vertex at one end. Since $\theta p = q$ and since θ carries chains free of fixed points into chains free of fixed points, we have

$$(1) \qquad \theta E_p = E_q, \qquad \theta E_q = E_p.$$

6.2. *If G has an ambiguous θ with two fixed points $\alpha = \alpha(p, q)$ and x, then G has a semi-split*

$$(2) \qquad G = (E_p + \alpha) + (E_q + x).$$

Proof. First, every edge γ of G appears once on the right of (2). For, G is cyclicly connected, so that γ lies on a circuit C containing α. C can contain at most two fixed points α and x. If γ is neither of these, then the subchain of C joining γ to an end of α and not containing x will be free of fixed points. Therefore such a γ lies in either E_p or E_q.

E_p and E_q have no common edges. For suppose β were in both E_p and E_q. Then, by (1), $\theta\beta$ would be in both, and there would be chains A_β and $A_{\theta\beta}$ connecting β and $\theta\beta$ to p. $A_\beta + A_{\theta\beta}$ would necessarily contain a chain L free of fixed points and having ends r and s on β and $\theta\beta$ such that $\theta(r) = s$. If L contains an interior vertex t with $\theta t \epsilon L$, then replace L by the subchain joining t to θt. Eventually we would thus obtain a new chain L with no such interior vertices t and with ends r' and s' such that $\theta(r') = s'$. This means that $L + \theta L$ is a fixed circuit under θ, contrary to the Definition 6.1 which asserts that an ambiguous θ has no such fixed points.

We now know that each edge of G appears exactly once on the right of (2). But by definition of E_p two edges on a common vertex not a fixed point would both lie in the same component E_p or E_q. Therefore the only vertices common to E_p and E_q are fixed vertices. If the given fixed point x is a vertex, then the components $E_p + \alpha$ and E_q of (2) obviously have in common only the vertices x and q, so that (2) is a semi-split at x and q. If x is an edge $x = x(s, t)$, it cannot have both its ends in one of the subgraphs E_p or E_q, say in E_q, for then the two subgraphs of (2) would have but one vertex in common, so that this vertex would cut G, contrary to the cyclic connectivity of G. Hence an end s of x must lie in E_p while the other end is in E_q. In this case (2) is a semi-split at q and s.

The component $E_p + \alpha$ is least, for it contains only one edge α ending on the split vertex q. Therefore (2) is a least split of G unless E_p or E_q is a chain. But if E_p is a chain then $\theta E_p = E_q$ must also be a chain, so that G is simply a circuit. The following conclusions result:

6.3. *If G has an ambiguous θ, then G has a least split or is a circuit.*

6.4. *If θ is ambiguous on the edge α, then α is contained in no minimal component of G.* For a minimal component M would, by definition, be contained in one of the components $E_p + \alpha$, $E_q + x$ of (2). By symmetry, M is also contained in one of $E_q - \alpha$, $E_p + x$. Therefore M is in one of E_p, E_q, and hence outside α, as asserted.

6.5. For a given map of G, we must distinguish when a given ambiguous automorph is a rotation and when it is a reflection. Hence the following definitions: Any automorph σ is *extendable* in a given map of G if each c. d. boundary C of the map is carried into a circuit, σC, which is a c. d. boundary. This means that any topological homeomorphism which corresponds to the automorphism σ as in § 1.1 is extendable to the sphere (cf. Theorem 5 in § 2). Suppose a map is given in which every automorph of a group Σ is extendable. If θ is ambiguous on the edge α, then α is on two c. d. boundaries J and K, while $\theta(\alpha) = \alpha$. Because θ is extendable, θ carries J into one of J or K, and K into one of J or K. Define $\chi(\theta)$ as

$$\begin{aligned} \chi(\theta) &= +1 \quad \text{if} \quad \theta(J) = K,\ \theta(K) = J, \\ \chi(\theta) &= -1 \quad \text{if} \quad \theta(J) = J,\ \theta(K) = K. \end{aligned} \tag{3}$$

If θ is also ambiguous on an edge β, α and β must by (2) belong to both J and K, and $\chi(\theta)$ is the same for α as for β. We call $\chi(\theta)$ the *character* of θ in the given map.[18] One verifies readily the following natural consequence of the definition.

6.6. *If G has a map in which every automorph of a group Σ is extendable, and if $\theta \epsilon \Sigma$ is ambiguous on G, then any conjugate $\bar{\theta}$ of θ in Σ is also ambiguous on G, and $\chi(\bar{\theta}) = \chi(\theta)$.*

7. The regular mapping method. We now prove the following theorem by induction on the number of distinct least splits of G.

---

[18] If $\chi(\theta) = +1$, then θ, when extended to the sphere, is homeomorphic to a rotation of the sphere, while $\chi(\theta) = -1$ would mean that θ yields a "reflection" of the sphere. Hence χ is in fact the group character of Σ which is $+1$ for all σ's which preserve orientation on the sphere and -1 for all other σ's.

THEOREM A. *If G is a cyclicly connected graph, and if Σ is an admissible group of automorphisms of G, then G can be mapped on the sphere so that every automorphism of Σ is extendable* (cf. § 6. 5). *Furthermore, if $\theta_1, \theta_2, \cdots, \theta_g$ are ambiguous automorphs of Σ on G, no two of which are conjugate in Σ, and if, for $i = 1, \cdots, g$, each ϵ_i is a number ± 1, then a map in which every $\sigma \epsilon \Sigma$ is extendable can be so chosen that each $\chi(\theta_i) = \epsilon_i$, for $i = 1, \cdots, g$.*

This asserts that the characters of the ambiguous automorphs can be arbitrarily assigned, subject to the inevitable restriction of § 6. 6.

7. 1. *First case of the proof:* G has no least split. This means as in § 4. 31 that G is either triply connected or a branch graph. If G is triply connected, the result of § 2. 2 shows that every homeomorphism is extendable in any map. If G is a branch graph, it consists of k suspended chains, $A_1, A_2, \cdots, A_k$, with the same ends p and q. If $k \leqq 3$, any map of G suffices. If $k \geqq 4$, we have a multiple split. Since the given group Σ is admissible, no parallel permutation is allowed; hence any σ with $\sigma p = p$, $\sigma q = q$ must have $\sigma A_i = A_i$ for $i = 1, \cdots, k$. Therefore such a σ is the identity. The only possible $\sigma \neq 1$ must then have $\sigma p = q$, $\sigma q = p$. Any two such automorphs σ and τ would have a product which would be a parallel permutation, hence the identity, so that $\sigma = \tau$ and $\sigma^2 = 1$.

Suppose then that the branch graph G has just one σ with $\sigma p = q$, $\sigma q = p$. If $\sigma A_i = A_i$ holds for all i, any map of G is satisfactory. If, however, $\sigma A_i = A_j$ for some $i \neq j$, the sum $A_i + A_j$ will be a fixed circuit of σ. From the conditions for an admissible group our branch graph could not then contain three fixed edges or vertices, and so there are at most two of the suspended chains with $\sigma A_i = A_i$. By a suitable choice of notation G will consist of the suspended chains

$$A_1, \cdots, A_s, \quad A'_1, \cdots, A'_s, \quad B_1, \; B_2,$$

where either B_1 or B_2 may be absent, and where

$$(1) \qquad \sigma B_1 = B_1, \quad \sigma B_2 = B_2, \quad \sigma A_i = A'_i, \quad \sigma A'_i = A_i, \quad (i = 1, \cdots, s).$$

The map of this branch graph with the c. d. boundaries

$$(2) \qquad B_1 + A_1, \quad B_1 + A'_1, \quad B_2 + A_s, \quad B_2 + A'_s, \quad A_i + A_{i+1}, \quad A'_i + A'_{i+1},$$

for $i = 1, \cdots, s - 1$, will then be a map in which the only automorph σ is extendable.

Finally, suppose that G has ambiguous automorphs θ_i. Since G has no least split, this implies, as in § 6. 3, that G must be a circuit. Any map has this circuit twice as a c. d. boundary, and the desired conclusion as to the ambiguous θ's is trivial. Thus Theorem A is established in this case.

7. 2. Assume that Theorem A is true for all graphs with fewer than n distinct least splits. Consider a cyclicly connected graph G with an admissible group Σ and with n least splits. By § 4. 23, G must have some minimal component. A skeleton G' can be constructed by replacing all the homologs of any one suitable minimal component $M_0 = M$ (see § 4, (2)). The skeleton G' has fewer than n least splits, by Lemma 4. 42, and is cyclicly connected by § 4. 43. Furthermore the group Σ' on G' is admissible by Lemma 5. 4, so that the induction assumption of our Theorem A applies to show that G' has a map in which all automorphs of Σ' are extendable. In this map, the new edge $\mu(p, q)$ used to replace the minimal component M_0 will appear in two c. d. boundaries, which may be written as

$$J = \mu + L, \qquad K = \mu + N, \tag{3}$$

where L and N are two chains in G' with the same ends, p and q, as $\mu = \mu_0$.

Lemma 7. 21. *Σ contains no σ of any order twisting M_0.*

For[19] if σ twists M_0, then[20] $\sigma\mu_0 = \mu_0$ and σ leaves the ends p and q fixed. Therefore σ, when suitably extended to the sphere from the map of G', has all the points of μ_0 as fixed points, and hence by Eilenberg's results must have a circuit of fixed points and must therefore be homeomorphic to a reflection of the sphere. This extended σ, and hence σ itself, is of order 2. The group Σ then contains an automorph σ of order 2 twisting M_0, contrary to the first condition for admissibility. The Lemma is established.

We next map $M + \alpha$, where $\alpha(p, q)$ is a new edge with the same ends as M_0. The chief difficulties arise from automorphs ρ in Σ' such that

$$\rho\mu = \mu, \qquad \rho \neq 1, \qquad \rho p = q, \qquad \rho q = p. \tag{4}$$

By Lemma 7. 21, there is at most one such ρ. But $\rho^2(M) = M$ and $\rho^2(p) = p$, while M has no twists; hence $\rho^2 = 1$.

If ρ is present, then the group generated by ρ is admissible on $M + \alpha$. For no twists or parallel permutations are present. The complement $G - M$

---

[19] This lemma could also be proven by the remark at the end of § 5.

[20] Henceforth we omit the primes distinguishing the automorphs σ in Σ from the σ' in Σ'.

contains a chain L with ends p and ρp. One argues then, as in § 6, that $L + \rho L$ contains either a fixed edge, vertex or circuit. Therefore, the fixed edge α of $M + \alpha$ corresponds to at least one fixed point of G under ρ. Every fixed point of $M + \alpha$ yields at least one fixed point of G. Thence one concludes that ρ, which is admissible on G, must be admissible on $M + \alpha$.

By § 7.1 there is a map of $M + \alpha$ in which ρ, if present, is extendable. Denote the c. d. boundaries of this map by

$$A + \alpha, \qquad B + \alpha, \qquad C_1, \cdots, C_s, \tag{5}$$

where only $A + \alpha$ and $B + \alpha$ contain α. Here the chains A and B can intersect[21] only at their ends p and q. Because ρ is extendable, it preserves the c. d. boundaries (5), so that ρA is A or B. This yields the alternatives

$$(6_1) \qquad \rho A = B, \ \rho B = A, \quad A + B \text{ a fixed circuit of } \rho,$$
$$(6_2) \qquad \rho A = A, \ \rho B = B, \quad A + B \text{ contains two fixed points of } \rho;$$

for in the second alternative A and B are transformed end for end and so each contains a fixed edge or vertex.

On the other hand, since ρ preserves the c. d. boundaries (3) of the map of G', we must have one of the following:

$$(7_1) \qquad \rho J = K, \qquad \rho K = J, \qquad \rho L = N, \qquad \rho N = L,$$
$$(7_2) \qquad \rho J = J, \qquad \rho K = K, \qquad \rho L = L, \qquad \rho N = N.$$

7.3. LEMMA. *If ρ of (4) is present, then either (6_1) and (7_1) hold together, or both (6_2) and (7_2) hold; provided that, if ρ is ambiguous in G', the map of G' is so chosen that $\chi(\rho)$ has a prescribed value* ± 1.

Proof. Suppose first that in the given map of G' there exists a fixed edge or vertex x contained in both L and N. On the map of G' let the point X be x or the midpoint of x and let the point Y be the midpoint of the other fixed edge μ. Join X to Y by two arcs β and γ lying in the domains bounded respectively by J and K. It follows that $\beta + \gamma$ is a circuit cutting G' into two pieces, one inside $\beta + \gamma$ and containing p, and one outside and containing q. But $\beta + \gamma$ meets G' only at the vertices X and Y invariant under ρ. Because $\rho p = q$, ρ must interchange the two pieces of G' which are cut apart at X and Y. Therefore, a combinatorial fixed point of G' under ρ can be in neither of these pieces and so must be at either X or Y. Hence ρ is ambiguous on the edge μ in G' according to the definition of § 6.1.

---

[21] This fact was proved in Lemma 5.41, if M is least, while it is obvious, if M is a branch graph.

By the induction assumption the map of G' can therefore be chosen so that $\chi(\rho)$ is $+1$ or -1; that is, so that either (7_1) or (7_2) holds. If the choice is regulated by the presence of (6_1) or (6_2), we get the conclusion of our Lemma.

Consider now the other case when L and N have no fixed edges or vertices in common. Either (7_1) or (7_2) holds in the given map of G'. If (7_2) holds, L and N must each contain a fixed edge or vertex, so that we have two distinct fixed points x' and y'. By the argument of § 5. 42, ρ has corresponding fixed points x and y in $G - M$. If M should then, as in (6_1), contain the fixed circuit $z = A + B$ under ρ, we would have three fixed points x, y, and z under the automorph ρ of order 2, while x and y lie in the combinatorial side $G - M$ of z. This contradicts the condition (ii) of our Theorem I as restated in the definition of an admissible group. Hence (7_2) implies (6_2), as in the Lemma.

Next suppose that (7_1) holds. If L has a vertex t with $\rho t \epsilon L$, then $\rho t \neq t$, for otherwise ρt would be in $\rho L = N$, and thus would be a common fixed point as in the case previously treated. As in a previous argument we construct in L a subchain L^* (which may equal L), with ends r and s, such that $\rho r = s$ while L^* contains no interior vertex t with $\rho t \epsilon L^*$. L^* cannot be a single fixed edge, for such an edge would be a common fixed point of L and N. Therefore $L^* + \rho L^*$ is a circuit which is a combinatorial fixed point under ρ. If, as in (6_2), M should then contain two fixed points y and z under ρ, we would again find a contradiction to the definition of admissibility. Thus (6_2) is impossible, so (7_1) implies (6_1), as required. The Lemma is now established.

7. 4. We now combine the maps of G' and $M = M_0$. Since all k of the minimal components M_i are homologous, there must be automorphs

$$\tau_0 = 1, \tau_1, \tau_2, \cdots, \tau_{k-1}$$

in the group Σ such that

$$(8) \qquad \tau_i \mu_0 = \mu_i, \qquad \tau_i M_0 = M_i, \qquad (i = 0, 1, \cdots, k-1).$$

Any σ in the group carries M_0 to some M_i, so that $\tau_i^{-1}\sigma$ sends M_0 to M_0. Hence any σ has one of the forms

$$(9) \qquad \sigma = \tau_i, \qquad \sigma = \tau_i \rho,$$

and therefore either the automorphs τ_i or these automorphs together with ρ generate the whole group.

In the map of the skeleton G', $\mu = \mu_0$ is on two c. d. boundaries J and K. Introduce the notation

$$J_i = \tau_i J, \qquad K_i = \tau_i K, \qquad A_i = \tau_i A, \qquad B_i = \tau_i B, \qquad (i = 1, \cdots, k-1). \tag{10}$$

Obviously, μ_i is on the boundaries J_i and K_i in the map of G'. Furthermore any τ_i applied to the map of $M_0 + \alpha$ in (5) gives a map of the corresponding $M_i + \alpha_i$ with the c. d. boundaries

$$A_i + \alpha_i, \qquad B_i + \alpha_i, \qquad \tau_i C_1, \cdots, \tau_i C_s, \tag{11}$$

where α_i is a new edge with the same ends as M_i. Maps of G can be constructed by superimposing on the map of the skeleton these maps of the minimal components. By Lemma 2.4 this can be done so that μ_i, formerly on the boundaries J_i and K_i, is so replaced by M_i that the chain A_i on the "rim" of M_i falls into the domain bounded by J_i, while the other chain B_i falls into the domain bounded by K_i. In other words, each boundary C of the map of G' yields in the map of G a boundary C^*, obtained from C by replacing each edge μ_i in C as follows:

$$\text{Replace } \mu_i \text{ by } A_i \text{ if } C = J_i; \text{ by } B_i \text{ if } C = K_i. \tag{12}$$

These circuits C^*, together with the circuits $\tau_i C_j$, are the c. d. boundaries of the map of G. This rule (12) is unambiguous except when $J_i = K_i$ for some i. This can happen only when G' is a circuit. Hence *assume for the present that G' is not a circuit.* (See, however, § 8.)

Since ρ leaves fixed the original set (5) of boundaries C_i, it follows readily from (9) that any σ of the group will carry a c. d. boundary of the form $\tau_i C_j$ into another circuit of the same form. It remains to show that the circuits $\sigma(C^*)$ are also c. d. boundaries. We shall prove more explicitly that

$$\sigma(C^*) = (\sigma C)^*, \tag{13}$$

where $(\sigma C)^*$ is certainly a c. d. boundary. For (13) it suffices to show that when μ_i is replaced by A_i in the construction (12), then $\sigma\mu_i$ is replaced by σA_i, and similarly for B_i. But for given σ and i there is a j so that $\sigma\mu_i = \mu_j$. For this σ, i, and j we need only show that one of the alternatives

$$\sigma A_i = A_j \quad \text{and} \quad \sigma J_i = J_j, \tag{14$_1$}$$

$$\sigma A_i = B_j \quad \text{and} \quad \sigma J_i = K_j \tag{14$_2$}$$

must hold. We know already that $\sigma\mu_i = \mu_j$, $\sigma M_i = M_j$ and hence by (9), (10), (6), and (7) that σA_i is either A_j or B_j and that σJ_i is either J_j or K_j.

Case 1. The group does not contain a ρ. Then $\sigma\tau_i$ carries μ_0 to μ_i to μ_j, hence $\sigma\tau_i = \tau_j$. Consequently,

$$\begin{aligned} \sigma A_i &= \sigma\tau_i A_0 = \tau_j A_0 = A_j, \\ \sigma J_i &= \sigma\tau_i J_0 = \tau_j J_0 = J_j, \end{aligned} \tag{15}$$

by the choice of notation in (10). The first alternative (14_1) holds, as required.

Case 2. Both (6_1) and (7_1) hold, as in Lemma 7.3. Then

$$\rho A = B, \qquad \rho B = A, \qquad \rho J = K, \qquad \rho K = J. \tag{16}$$

Now consider (14) for any σ and i. Since $\sigma\tau_i$ carries μ_0 to μ_j, $\sigma\tau_i$ is either τ_j or $\tau_j\rho$ by (9). If $\sigma\tau_i = \tau_j$, then (15) holds again, as required in (14). Alternatively, if $\sigma\tau_i = \tau_j\rho$ holds, then (16) and (10) entail

$$\begin{aligned} \sigma A_i &= \sigma\tau_i A_0 = \tau_j\rho A_0 = \tau_j B_0 = B_j, \\ \sigma J_i &= \sigma\tau_i J_0 = \tau_j\rho J_0 = \tau_j K_0 = K_j, \end{aligned}$$

which is the second alternative of (14).

Case 3. Both (6_2) and (7_2) hold, as in Lemma 7.3. Then

$$\rho A = A, \qquad \rho B = B, \qquad \rho J = J, \qquad \rho K = K.$$

From this point we obtain the result (14) much as in the previous case.

Thus in all cases we have established one of (14_1) and (14_2), so that the c. d. boundaries C^* of our map of G introduced in (12) are in fact carried into boundaries by any σ. Thus all the automorphs σ of the group Σ are extendable.

7.5. Now let Σ contain automorphs $\theta_1, \theta_2, \cdots, \theta_g$, no two of them conjugate, and each θ_i ambiguous on an edge β_i of G, and let g numbers $\epsilon_i = \pm 1$ be given. By § 6.4, each β_i is also an edge of the skeleton G', so that each θ_i is ambiguous on the corresponding β_i in G'. Hence our induction assumption yields a map of G' in which $\chi(\theta_i) = \epsilon_i$ for $i = 1, \cdots, g$. If the "reflection" ρ should be present and ambiguous on G', then $\rho(M_0) = M_0$ implies that ρ is not ambiguous on G, so that ρ cannot be conjugate to any of the ambiguous θ's (see § 6.6). Thus the map of G' with $\chi(\theta_i) = \epsilon_i$ can also be chosen so that $\chi(\rho)$ has an assigned value, and hence so that Lemma 7.3 applies. Construct the corresponding map of G as in (12). The definition of χ and the relation (13) together imply that $\chi(\theta_i)$ has the same value in this map of G as it does in the map of G'. Thus we have an "extendable" map of G with the assigned characters $\chi(\theta_i) = \epsilon_i$, and the induction is complete in this case.

8. The circular case. It remains only to consider the omitted case of § 7.4 in which the skeleton G' is a circuit. Choose the notation so that the new edges are $\mu_i = \mu_i(p_i, q_i)$, where[22] the vertices $p_i\, q_i$ are in the cyclic order

$$p_0 q_0 p_1 q_1 p_2 q_2 \cdots p_{k-1} q_{k-1} \tag{1}$$

---

[22] Throughout this section, subscripts i, j, etc. run from 0 to $k-1$, or may be reduced mod k.

on the circuit G'. In addition to the μ_i, G' contains chains $\nu_i = \nu_i(q_i, p_{i+1})$. Any automorph carries μ's to μ's, hence carries ν's to ν's, so that without altering Σ we can assume that each ν_i is a single edge.[23] Thus G' has the edges $\mu_0, \nu_0, \mu_1, \nu_1, \cdots, \mu_{k-1}, \nu_{k-1}$ in that order, and

$$G = M_0 + \nu_0 + M_1 + \nu_1 + \cdots + M_{k-1} + \nu_{k-1}, \qquad \nu_i = \nu_i(q_i, p_{i+1}),$$

where the subgraphs on the right have no common edges and only the vertices (1), as indicated, in common.

Every automorph σ of G' carries each μ_i into some μ_j and preserves or inverts the order (1). Hence every σ belongs to the dihedral group generated by ω and ρ, where ω is a "rotation" of $2\pi/k$ radians and ρ a "reflection" in the "diameter" through ν_0:

$$(2) \qquad \omega p_i = p_{i+1}, \quad \omega q_i = q_{i+1} \quad \omega\mu_i = \mu_{i+1}, \quad \omega\nu_i = \nu_{i+1};$$

$$(3) \qquad \rho p_i = q_{k+1-i}, \quad \rho q_i = p_{k+1-i}, \quad \rho\mu_i = \mu_{k+1-i}, \quad \rho\nu_i = \nu_{k-i}.$$

One readily obtains the following relations:

$$(4) \qquad \omega^k = 1, \quad \rho^2 = 1, \quad \rho\omega^i = \omega^{-i}\rho, \quad \omega^i\rho = \rho\omega^{-i}.$$

The group Σ' must contain automorphs taking μ_0 to each μ_i. The only σ's with $\sigma(\mu_0) = \mu_1$ are $\sigma = \rho$ and $\sigma = \omega$. If ω is absent and $k > 2$, then the only σ's with $\sigma(\mu_0) = \mu_2$ are ω^2 and $\rho\omega^{k-1} = \omega\rho$. But $\omega\rho$ and ρ together give ω. Hence Σ' must contain ω or else must contain ρ and ω^2. The only such groups Σ' are the three following ones, where each group is determined by the generators given in braces:

$$(5) \qquad \Sigma_1\colon \{\omega^2, \rho\}, \text{ with } k = 2t; \quad \Sigma_2\colon \{\omega\}; \quad \Sigma_3\colon \{\omega, \rho\}.$$

For Σ_3, note that $\rho\omega(\mu_0) = \mu_0$, so that $\rho\omega$ leaves μ_0 fixed.

To map G, first take a new edge $\alpha(p_0, q_0)$ and choose a map of $M_0 + \alpha$ as in § 7, (5). If the group is Σ_3, then $\rho\omega$ is an automorph of $M + \alpha$ if we set $\rho\omega(\alpha) = \alpha$. In this case the map of $M + \alpha$ is to be chosen, as in § 7. 2, so that $\rho\omega$ is extendable on the map. Choose automorphs $\tau_i \in \Sigma'$ such that $\tau_i\mu_0 = \mu_i$; if ω is present, choose $\tau_i = \omega^i$ and apply § 7, (10). Then $M_i + \alpha_i$, where $\alpha_i = \alpha_i(p_i, q_i)$, has a map with the c. d. boundaries of § 7, (11). Superimpose all these maps on G' so that the A's all fall on one side of G'; that is, so that G has a map with the c. d. boundaries, $\tau_i C_j$, for $i = 0, \cdots, k-1$, and $j = 1, \cdots, s$, and also two boundaries

$$(6) \qquad D_1 = A_0 + \nu_0 + A_1 + \nu_1 + \cdots + A_{k-1} + \nu_{k-1},$$

$$(7) \qquad E_1 = B_0 + \nu_0 + B_1 + \nu_1 + \cdots + B_{k-1} + \nu_{k-1}.$$

---

[23] If $q_i = p_{i+1}$, ν_i is a single vertex. The formulas below will apply also to this case.

If the group is Σ_1 or Σ_2, the τ_i include the whole group, and every τ_i is extendable, so that we have a map with all automorphs extendable.

If the group is Σ_3, $\tau_i = \omega^i$, $\rho\omega(\mu_0) = \mu_0$, and $\rho\omega$ is extendable on $M_0 + \alpha$, so that $\rho\omega(A_0)$ is A_0 or B_0. Therefore one of the alternatives

$$(8_1) \qquad \rho A_0 = A_1, \qquad \rho B_0 = B_1;$$

$$(8_2) \qquad \rho A_0 = B_1, \qquad \rho B_0 = A_1;$$

must hold.[24] In the first alternative, (4) and the definition $\tau_i = \omega^i$ show that

$$(9_1) \qquad \rho A_i = \rho\omega^i A_0 = \omega^{-i}\rho A_0 = \omega^{-i}A_1 = A_{1-i}, \qquad \rho B_i = B_{1-i}.$$

In the second alternative, one similarly calculates that

$$(9_2) \qquad \rho A_i = B_{1-i}, \qquad \rho B_i = A_{1-i}.$$

It follows that ρ interchanges D_1 and E_1 or leaves them each fixed according as (8_2) or (8_1) holds. But $\rho\omega$ also leaves the set of all C_s fixed, because the map of $M_0 + \alpha$ was picked so that $\rho\omega$ is extendable on this map. Hence $\rho(\tau_i C_j) = \omega^{1-i}(\rho\omega C_j)$ is a c. d. boundary, and ρ is extendable in the map of G. But ω is extendable by the construction of (6) and (7), so all σ's in Σ_3 are extendable.

An automorph θ of G could be ambiguous only on some edge ν_j, by § 6. 4. But $\rho\omega^{k-2j}$ is the only $\sigma \neq 1$ with ν_j as a fixed edge. Hence the ambiguous θ's are all in the set

$$(10) \qquad \theta_j = \rho\omega^{k-2j}, \qquad (j = 0, 1, \cdots, k-1).$$

Every $\theta_j^2 = 1$. If k is odd, then $k = 2t + 1$, $\theta_j(\mu_{t+j+1}) = \mu_{t+j+1}$ by (2) and (3), and θ_j interchanges the ends of M_{t+j+1}. As in Lemma 5. 41, θ_j must then have a fixed circuit or two fixed points in M_{t+j+1}. These, in addition to the fixed point ν_j, show that θ_j is not ambiguous on G. There remains only the case $k = 2t$, in which event θ_j is ambiguous on both ν_j and ν_{j+t} in G.

Possible conjugates of θ_j are $\omega^i\theta_j\omega^{-i} = \theta_{i+j}$ and $\rho\theta_j\rho^{-1} = \theta_{t-j}$. Therefore all θ's are conjugate in Σ_3, while in Σ_2 of (5) there are no θ's present. For the group Σ_1 of (5), θ_0 has conjugates $\theta_0, \theta_2, \theta_4, \cdots$ and $\theta_t, \theta_{t-2}, \cdots$, where $k = 2t$. Thus all θ's are conjugates if t is odd, while if t is even θ_0 and θ_1 are not conjugates, but every θ is conjugate to one of θ_0 or θ_1.

For alternative maps when $k = 2t$, we again superimpose the maps of $M_i + \alpha_i$ on G', but now so that the c. d. boundaries are not (6) and (7), but

$$(11) \qquad \begin{aligned} D_2 &= \Sigma\nu_i + A_0 + B_1 + A_2 + B_3 + \cdots + A_{k-2} + B_{k-1}, \\ E_2 &= \Sigma\nu_i + B_0 + A_1 + B_2 + A_3 + \cdots + B_{k-2} + A_{k-1}, \end{aligned}$$

---

[24] The automorph ρ, though a "reflection" of the circuit G', is not necessarily a "reflection" of G.

where $\Sigma\nu_i$ is the subgraph of all the ν_i. If $k = 2t$, and t is even, use

$$\begin{aligned} D_3 &= \Sigma\nu_i + A_0 + A_1 + B_2 + B_3 + A_4 + A_5 + B_6 + B_7 + \cdots + B_{k-2} + B_{k-1}, \\ E_3 &= \Sigma\nu_i + B_0 + B_1 + A_2 + A_3 + B_4 + B_5 + A_6 + A_7 + \cdots + A_{k-2} + A_{k-1}, \end{aligned} \tag{12}$$

as the c. d. boundaries, or else use the circuits D_4 and E_4 obtained from D_3 and E_3 by increasing all subscripts on the right of (12) by 1. Call the latter map (12 +).

If the group is Σ_3, and $k = 2t$, then ω is extendable in either map (7) or map (11). All θ's in (10) are conjugate to $\theta_0 = \rho$. As in (9), ρ carries μ_i to μ_{1-i}, hence takes odd subscripts to even ones. If (9_1) holds, ρ carries A's to A's, so that $\rho(D_2) = E_2$, $\rho(E_2) = D_2$. In both maps ρ is extendable, but, by the definition in § 6. 5, $\chi(\rho)$ is $+1$ in (11), and -1 in (7). If the other alternative (9_2) holds, ρ is again extendable, but $\chi(\rho) = +1$ in (7), and -1 in (11). In either event a map can be picked so that all σ's are extendable, and ρ, the typical ambiguous automorph, has either character desired.

If the group is Σ_1, and $k = 2t$, similar calculations can be made. By (10), (2), and (3), one has

$$\rho\mu_{2i} = \mu_{1-2i}, \qquad \rho\mu_{2i+1} = \mu_{-2i}, \qquad \theta_1\mu_{2i} = \mu_{3-2i}, \qquad \theta_1\mu_{2i+1} = \mu_{2-2i}.$$

If t is odd, all θ's are conjugate to $\theta_0 = \rho$, and (7) and (11) again give maps with ω^2 and ρ extendable and with $\chi(\rho) = -1$ in (7), and $\chi(\rho) = +1$ in (11). If t is even, all θ's are conjugate to θ_0 or θ_1. But ω^2 and ρ are extendable in all four maps, and the characters $\chi(\rho)$ and $\chi(\theta_1)$ are, in that order, $(-1, -1)$ in (7), $(+1, +1)$ in (11), $(-1, +1)$ in (12), $(+1, -1)$ in (12 +). This gives all four possible sets of characters.

This completes the proof that when the skeleton is a circuit there is a map of G in which all σ's are extendable and in which the characters of ambiguous automorphs are arbitrarily assigned. The induction for Theorem A is finished, and with it the proof of Theorem 1 and its consequences in § 1.

Cornell University,
University of Arkansas.

BIBLIOGRAPHY.

Adkisson, V. W.

1. "Cyclicly connected continuous curves whose complementary domain boundaries are homeomorphic," *Comptes Rendues des séances de la Société des Sciences et des Lettres de Varsovie*, vol. 23 (1930), Classe III.

2. "On extending a continuous (1 — 1) correspondence of continuous curves on a sphere," *Comptes Rendues des séances de la Société des Sciences et des Lettres de Varsovie*, vol. 27 (1934), Classe III.

Adkisson, V. W., and MacLane, S.

1. "Planar graphs whose homeomorphisms can all be extended for any mapping on the sphere," *American Journal of Mathematics*, vol. 59 (1937), pp. 823-832.

Eilenberg, S.

1. "Sur les transformations périodique de la surface de sphère," *Fundamenta Mathematicae*, vol. 22 (1934), pp. 28-41.

Gehman, H. M.

1. "On extending a continuous (1 — 1) correspondence of two plane continuous curves to a correspondence of their planes," *Transactions of the American Mathematical Society*, vol. 28 (1926), pp. 252-265.

2. "On extending a continuous (1 — 1) correspondence of two plane continuous curves to a correspondence of their planes," second paper, *Transactions of the American Mathematical Society*, vol. 31 (1929), pp. 241-252.

3. "On extending a homeomorphism between two subsets of spheres," *Bulletin of the American Mathematical Society*, vol. 42 (1936), pp. 79-81.

MacLane, S.

1. "Some unique separation theorems for graphs," *American Journal of Mathematics*, vol. 57 (1935), pp. 805-820.

2. "A combinatorial condition for planar graphs," *Fundamenta Mathematicae*, vol. 28 (1937), pp. 22-32.

3. "A structural characterization of combinatorial planar graphs," *Duke Mathematical Journal*, vol. 3 (1937), pp. 460-472.

Moore, R. L.

1. "Conditions under which one of two given closed linear point sets may be thrown into the other one by a continuous transformation of the plane into itself," *American Journal of Mathematics*, vol. 48 (1926), pp. 67-72.

Whitney, H.

1. "Non-separable and planar graphs," *Transactions of the American Mathematical Society*, vol. 34 (1932), pp. 339-362.

2. "Planar graphs," *Fundamenta Mathematicae*, vol. 21 (1933), pp. 73-84.

3. "Congruent graphs and the connectivity of graphs," *American Journal of Mathematics*, vol. 54 (1932), pp. 150-168.

4. "On the classification of graphs," *American Journal of Mathematics*, vol. 55 (1933), pp. 236-244.

Whyburn, G. T.

1. "Cyclicly connected continuous curves," *Proceedings of the National Academy of Sciences*, vol. 13 (1927), pp. 31-38.

2. "Concerning the structure of a continuous curve," *American Journal of Mathematics*, vol. 50 (1928), pp. 167-194

3. "On the structure of continua," *Bulletin of the American Mathematical Society*, vol. 42 (1936), pp. 49-73.

Whyburn, G. T. and Kuratowski, C.

1. "Sur les éléments cycliques et leurs applications," *Fundamenta Mathematicae*, vol. 16 (1930), pp. 305-331.

ASYMPTOTIC CURVES ON A SURFACE.*

By E. P. Lane and M. L. MacQueen.

1. Introduction. The purpose of this paper is to make some contributions to the theory of the asymptotic curves on an analytic non-ruled surface in ordinary projective space. Section 2 contains a summary of portions of the classical analytical theory of the projective differential geometry of curves and surfaces which are used in later developments. Power series expansions in non-homogeneous projective coördinates for the parametric asymptotic curves on the surface are computed to terms of higher degree than have hitherto been considered. By means of these power series a brief study is made of the asymptotic curves regarded as space curves.

2. Analytic basis. If the four homogeneous projective coördinates x of a variable point on an analytic non-ruled surface S in ordinary space are given as analytic functions of two independent variables u, v, and if the parametric net on S is the asymptotic net, then the functions x are solutions of a system of differential equations which can be reduced to the form

$$\begin{aligned} x_{uu} &= px + \theta_u x_u + \beta x_v, \\ x_{vv} &= qx + \gamma x_u + \theta_v x_v \end{aligned} \qquad (\theta = \log \beta\gamma). \tag{1}$$

The coefficients of these equations are functions of u, v and satisfy certain integrability conditions.

The formulas for the third derivatives of x expressed as linear combinations of x, x_u, x_v, x_{uv} are found from system (1) to be

$$\begin{aligned} x_{uuu} &= (p_u + p\theta_u)x + (p + \theta_u^2 + \theta_{uu})x_u + (\beta_u + \beta\theta_u)x_v + \beta x_{uv}, \\ x_{uuv} &= (p_v + \beta q)x + kx_u + \pi x_v + \theta_u x_{uv}, \\ x_{uvv} &= (q_u + \gamma p)x + \chi x_u + kx_v + \theta_v x_{uv}, \\ x_{vvv} &= (q_v + q\theta_v)x + (\gamma_v + \gamma\theta_v)x_u + (q + \theta_v^2 + \theta_{vv})x_v + \gamma x_{uv}, \end{aligned} \tag{2}$$

where π, χ, k are defined by the formulas

$$\pi = p + \beta\psi, \qquad \chi = q + \gamma\phi, \qquad k = \beta\gamma + \theta_{uv}, \tag{3}$$

and ϕ, ψ by

$$\phi = (\log \beta\gamma^2)_u, \qquad \psi = (\log \beta^2\gamma)_v. \tag{4}$$

---

* Received March 3, 1938.

Furthermore, we find

$$x_{uuuu} = A_4 x + B_4 x_u + C_4 x_v + D_4 x_{uv}, \tag{5}$$

where

$$\begin{aligned} A_4 &= (p_u + p\theta_u)_u + p(p + \theta_u^2 + \theta_{uu}) + \beta(p_v + \beta q), \\ B_4 &= (p_u + p\theta_u) + \theta_u(p + \theta_u^2 + \theta_{uu}) + (p + \theta_u^2 + \theta_{uu})_u + \beta k, \\ C_4 &= \beta(p + \theta_u^2 + \theta_{uu}) + (\beta_u + \beta\theta_u)_u + \beta\pi, \\ D_4 &= 2(\beta_u + \beta\theta_u). \end{aligned} \tag{6}$$

In fact it is possible to express every u-derivative of x as a linear combination of x, x_u, x_v, x_{uv} by an equation of the form

$$\partial x^i/\partial u^i = A_i x + B_i x_u + C_i x_v + D_i x_{uv}, \tag{7}$$

where

$$\begin{aligned} A_i &= A_{i-1,u} + pB_{i-1} + (p_v + \beta q)D_{i-1}, \\ B_i &= A_{i-1} + \theta_u B_{i-1} + B_{i-1,u} + kD_{i-1}, \\ C_i &= \beta B_{i-1} + C_{i-1,u} + \pi D_{i-1}, \\ D_i &= C_{i-1} + \theta_u D_{i-1} + D_{i-1,u}. \end{aligned} \tag{8}$$

The parametric vector equation of an analytic curve is

$$x = x(t), \tag{9}$$

the parameter being t. These coördinates x satisfy an ordinary differential equation of the form

$$x^{iv} + 4p_1 x''' + 6p_2 x'' + 4p_3 x' + p_4 x = 0 \qquad (x' = dx/dt, \cdots), \tag{10}$$

the coefficients being functions of t. Let P_2, P_3, P_4 be defined by the formulas

$$\begin{aligned} P_2 &= p_2 - p_1^2 - p_1', \\ P_3 &= p_3 - 3p_1p_2 + 2p_1^3 - p_1'', \\ P_4 &= p_4 - 4p_1p_3 - 3p_2^2 + 12p_1^2p_2 - 6p_1^4 - p_1'''. \end{aligned} \tag{11}$$

Then two invariants θ_3, θ_4 of the differential equation are defined by the formulas

$$\theta_3 = P_3 - 3/2\,P_2', \quad \theta_4 = P_4 - 81/25\,P_2^2 - 2P_3' + 6/5\,P_2''. \tag{12}$$

If the differential equation of the form (10) for the asymptotic u-curves on the surface is calculated by means of equations (1), (2), (5), then the coefficients of this equation are found to be given by the formulas

$$
(13)\quad
\begin{aligned}
2p_1 &= \phi - 3\theta_u,\\
6p_2 &= \phi^2 + \phi_u - 7\theta_u\phi - 4\theta_{uu} + 11\theta_u^2 - 2p - \beta\psi,\\
4p_3 &= 7\theta_u\theta_{uu} - 6\theta_u^3 - \theta_{uuu} + 6p\theta_u - 2p_u + 5\theta_u^2\phi\\
&\qquad - \theta_u\phi^2 - 2p\phi - 2\theta_{uu}\phi - \theta_u\phi_u + \theta_u\beta\psi - \beta k,\\
p_4 &= p^2 + p\beta\psi - 6p\theta_u^2 + 2p\theta_{uu} - p\phi_u + 5p\theta_u\phi\\
&\qquad + 5p_u\theta_u - p_{uu} - p\phi^2 - 2p_u\phi - \beta(p_v + \beta q).
\end{aligned}
$$

The invariants θ_3, θ_4 defined by the formulas (12) can be calculated for the u-curves, but we shall omit the writing of these results here.

3. Power series expansions for the asymptotic curves. In ordinary space an analytic curve can be defined by expressing two of the non-homogeneous coördinates of a point on the curve as power series in the third coördinate. Power series expansions for the u-curve can be calculated in the following way.

The coördinates of any point X near the point x and on the u-curve can be represented by Taylor's expansion as power series in the increment Δ_u corresponding to displacement on S from x to the point X along the curve:

$$
(14)\qquad X = x + x_u\Delta u + \tfrac{1}{2}x_{uu}\Delta u^2 + \tfrac{1}{6}x_{uuu}\Delta u^3 + \tfrac{1}{24}x_{uuuu}\Delta u^4 + \cdots.
$$

Expressing each of $x_{uu}, x_{uuu}, \cdots$ as a linear combination of x, x_u, x_v, x_{uv}, we find

$$
X = x_1 x + x_2 x_u + x_3 x_v + x_4 x_{uv},
$$

where the local coördinates $x_1, \cdots, x_4$ of the point X are given by the expansions

$$
(15)\quad
\begin{aligned}
x_1 &= 1 + \tfrac{1}{2}p\Delta_u^2 + \tfrac{1}{6}(p_u + p\theta_u)\Delta_u^3 + \tfrac{1}{24}A_4\Delta_u^4 + \cdots,\\
x_2 &= \Delta u + \tfrac{1}{2}\theta_u\Delta u^3 + \tfrac{1}{6}(p + \theta_u^2 + \theta_{uu})\Delta u^3\\
&\qquad + \tfrac{1}{24}B_4\Delta_u^4 + \tfrac{1}{120}B_5\Delta_u^5 + \cdots,\\
x_3 &= \tfrac{1}{2}\beta\Delta u^2 + \tfrac{1}{6}(\beta_u + \beta\theta_u)\Delta u^3 + \tfrac{1}{24}C_4\Delta u^4\\
&\qquad + \tfrac{1}{120}C_5\Delta u^5 + \tfrac{1}{720}C_6\Delta u^6 + \cdots,\\
x_4 &= \tfrac{1}{6}\beta\Delta u^3 + \tfrac{1}{12}(\beta_u + \beta\theta_u)\Delta u^4 + \tfrac{1}{120}D_5\Delta u^5\\
&\qquad + \tfrac{1}{720}D_6\Delta u^6 + \tfrac{1}{5040}D_7\Delta u^7 + \cdots.
\end{aligned}
$$

Introducing non-homogeneous coördinates by the definitions

$$
(16)\qquad x = x_2/x_1,\quad y = x_3/x_1,\quad z = x_4/x_1,
$$

we find

$$
\begin{aligned}
x &= \Delta u + \tfrac{1}{2}\theta_u \Delta u + \tfrac{1}{6}(\theta_u{}^2 + \theta_{uu} - 2p)\Delta u^3 \\
&\quad + \tfrac{1}{24}(B_4 - 4p_u - 10p\theta_u)\Delta_u^4 \\
&\quad + \tfrac{1}{120}[B_5 - 5A_4 - 10\theta_u(p_u + p\theta_u)]\Delta u^5 + \cdots, \\
y &= \tfrac{1}{2}\beta\Delta u^2 + \tfrac{1}{6}(\beta_u + \beta\theta_u)\Delta u^3 + \tfrac{1}{24}(C_4 - 6\beta p)\Delta u^4 \\
&\quad + \tfrac{1}{120}[C_5 - 10\beta(p_u + p\theta_u) - 10p(\beta_u + \beta\theta_u)]\Delta u^5 \\
&\quad + \tfrac{1}{720}[C_6 - 15\beta A_4 - 20(\beta_u + \beta\theta_u)(p_u + p\theta_u) \\
&\qquad - 15p(C_4 - 6\beta p)]\Delta u^6 + \cdots, \\
z &= \tfrac{1}{6}\beta\Delta u^3 + \tfrac{1}{12}(\beta_u + \beta\theta_u)\Delta u^4 + \tfrac{1}{120}(D_5 - 10\beta p)\Delta u^5 \\
&\quad + \tfrac{1}{720}[D_6 - 20\beta(p_u + p\theta_u) - 30p(\beta_u + \beta\theta_u)]\Delta u^6 \\
&\quad + \tfrac{1}{5040}[D_7 - 21p(D_6 - 10\beta p) - 35\beta A_4 \\
&\qquad - 70(\beta_u - \beta\theta_u)(p_u + p\theta_u)]\Delta u^7 + \cdots.
\end{aligned}
\tag{17}
$$

Inverting the first of these series we obtain

$$\Delta u = x + c_2 x^2 + c_3 x^3 + c_4 x^4 + c_5 x^5 + \cdots, \tag{18}$$

where

$$
\begin{aligned}
c_2 &= -\tfrac{1}{2}\theta_u, \qquad c_3 = -\tfrac{1}{6}(\theta_{uu} - 2\theta_u{}^2 - 2p), \\
c_4 &= -\tfrac{1}{24}(\theta_{uuu} - 7\theta_u\theta_{uuu} + 6\theta_u{}^3 + 12p\theta_u - 2p_u + \beta k), \\
c_5 &= -\tfrac{1}{120}\,(\theta_{uuuu} - 11\theta_u\theta_{uuu} + 46\theta_u{}^2\theta_{uu} - 24\theta_u{}^4 - 7\theta_{uu}{}^2 \\
&\quad + 24p(\theta_{uu} - 3\theta_u{}^2 - p) + 20p_u\theta_u - 2p_{uu} \\
&\quad + \beta[k_u + 6k\theta_u - 3k\phi - 4(p_v + \beta q)]).
\end{aligned}
\tag{19}
$$

If we substitute the series (18) for Δu in the last two of the series (17) we arrive at *the power series expansions for the asymptotic u-curve*, namely,

$$
\begin{aligned}
y &= \tfrac{1}{2}\beta x^2 - \tfrac{1}{6}\beta\phi x^3 + \tfrac{1}{24}\beta a_4 x^4 + \tfrac{1}{120}\beta a_5 x^5 + \tfrac{1}{720}\beta a_6 x^6 + \cdots, \\
z &= \tfrac{1}{6}\beta x^3 - \tfrac{1}{12}\beta\phi x^4 + \tfrac{1}{120}\beta b_5 x^5 \\
&\qquad + \tfrac{1}{720}\beta b_6 x^6 + \tfrac{1}{5040}\beta b_7 x^7 + \cdots,
\end{aligned}
\tag{20}
$$

where

$$
\begin{aligned}
a_4 &= \phi^2 + \theta_u\phi - \phi_u + 4p + \beta\psi, \\
a_5 &= a_{4u} - (2\theta_u + \phi)a_4 - 10p\phi - 2\beta\phi\psi - 4\beta k, \\
a_6 &= a_{5u} - (3\theta_u + \phi)a_5 + 3(6p + \beta\psi)a_4 \\
&\qquad + 20\beta\phi k - 2\pi\beta\psi + 10\beta(p_v + \beta q), \\
b_5 &= 3a_4 - 2\beta\psi, \\
b_6 &= 4a_5 + 10\beta\phi\psi + 2\beta(k + \phi_v - 3\theta_{uv}), \\
b_7 &= 5a_6 - (3\theta_u + \phi)(b_6 - 4a_5) \\
&\qquad + b_{6u} - 4a_{5u} - 4\beta\psi(b_5 + 10p) - 10\beta\phi k.
\end{aligned}
\tag{21}
$$

Analogous expansions for the v-curve can be written by making the appropriate symmetrical interchanges of the symbols.

4. A canonical form for the expansions. By suitable choice of the coördinate system the power series expansions (20) can be reduced to a simple canonical form which it is the purpose of this section to obtain.

First of all, *the osculating twisted cubic* at the point $O(0,0,0)$ of the u-curve defined by equations (20) is found to have the parametric equations

$$\begin{aligned} x_1 &= 1 + Ct^2 - Dt^3, & x_3 &= \tfrac{1}{2}\beta t^2, \\ x_2 &= t + \tfrac{1}{6}\phi t^2 + Ht^3, & x_4 &= \tfrac{1}{6}\beta t^3, \end{aligned} \tag{22}$$

wherein C, H, D are defined by the formulas

$$\begin{aligned} C &= \tfrac{1}{60}(3a_4 - 5\phi^2 - 12\beta\psi), \\ H &= -\tfrac{1}{180}(3a_4 - 5\phi^2 + 18\beta\psi), \\ D &= \tfrac{1}{60}a_5 + \tfrac{1}{12}\phi a_4 - \tfrac{2}{27}\phi^3. \end{aligned} \tag{23}$$

Incidentally, it will be observed that C, H satisfy the relation

$$C + 3H = -\tfrac{1}{2}\beta\psi. \tag{24}$$

The osculating twisted cubic may also be written in the form of two power series expansions which must agree with the series (20) up to and including the fifth powers of x. For this result we find

$$\begin{aligned} y &= \tfrac{1}{2}\beta x^2 - \tfrac{1}{6}\beta\phi x^3 + \tfrac{1}{24}\beta a_4 x^4 + \tfrac{1}{120}\beta a_5 x^5 + \tfrac{1}{720}\beta \bar{a}_6 x^6 + \cdots, \\ z &= \tfrac{1}{6}\beta x^3 - \tfrac{1}{12}\beta\phi x^4 + \tfrac{1}{120}\beta b_5 x^5 \\ &\qquad + \tfrac{1}{720}\beta \bar{b}_6 x^6 + \tfrac{1}{5040}\beta \bar{b}_7 x^7 + \cdots, \end{aligned} \tag{25}$$

where

$$\begin{aligned} \bar{a}_6 &= -7\phi a_5 - \tfrac{140}{9}\phi^4 - 140(C - 2H)\phi^2 \\ &\qquad + 360(2C^2 - 8CH + 7H^2), \\ \bar{b}_6 &= 4a_5 + 10\beta\phi\psi, \\ \bar{b}_7 &= 5a_6 - \tfrac{140}{3}\beta\phi^2\psi - 60\beta\psi(5C - 7H). \end{aligned} \tag{26}$$

Let us make the transformation

$$\begin{aligned} x &= (X + \tfrac{1}{3}\phi Y + HZ)/(1 + CY + DZ), \\ y &= (\tfrac{1}{2}\beta Y)/(1 + CY + DZ), \\ z &= (\tfrac{1}{6}\beta Z)/(1 + CY + DZ), \end{aligned} \tag{27}$$

from the old coördinates x, y, z to new coördinates X, Y, Z. This transformation moves the vertex $(0,0,0,1)$ of the tetrahedron of reference to a point on the osculating twisted cubic of the u-curve. Furthermore, the edge $X_1 = X_2 = 0$ of the new tetrahedron is tangent to the osculating cubic at the point $(0,0,0,1)$, and the face $X_1 = 0$ is the osculating plane of the cubic at this point. The unit point is somewhere on the osculating cubic whose equations become

$$Y = X^2, \qquad Z = X^3. \tag{28}$$

Moreover, we find that *the equations* (20) *of the u-curve can be written in the canonical form*

$$\begin{aligned} Y &= X^2 + A_6X^6 + \cdots, \\ Z &= X^3 + B_6X^6 + B_7X^7 + \cdots, \end{aligned} \tag{29}$$

the coefficients A_6, B_6, B_7 *being defined by the formulas*

$$\begin{aligned} A_6 &= \tfrac{1}{360}(a_6 - \bar{a}_6), \\ B_6 &= \tfrac{1}{720}(b_6 - \bar{b}_6), \\ B_7 &= \tfrac{1}{840}(b_7 - \bar{b}_7 + \phi B_6). \end{aligned} \tag{30}$$

If we replace $\bar{b}_6$ in the second of equations (30) by its value found in equations (26), we find

$$B_6 = \tfrac{1}{60}\beta[\beta\gamma - (\log\beta)_{uv}]. \tag{31}$$

The coefficient B_6 is found to be an invariant and differs from the invariant θ_3 of the differential equation of the u-curve only by a constant factor. In fact, if θ_3 is calculated, it is found that

$$B_6 = -\tfrac{1}{15}\theta_3. \tag{32}$$

Moreover, further calculation shows that

$$A_6 = -\tfrac{1}{180}(5\theta_4 + 2\theta'_3), \qquad B_7 = -\tfrac{1}{420}(25\theta_4 + 14\theta'_3), \tag{33}$$

so that we have

$$B_7 - 3A_6 = \tfrac{1}{42}\theta_4. \tag{34}$$

Therefore *the expansions for the u-curve may be written in the form*

$$\begin{aligned} Y &= X^2 - \tfrac{1}{180}(5\theta_4 + 2\theta'_3)X^6 + \cdots, \\ Z &= X^3 - \tfrac{1}{15}\theta_3X^6 - \tfrac{1}{420}(25\theta_4 + 14\theta'_3)X^7 + \cdots. \end{aligned} \tag{35}$$

5. Asymptotic curves regarded as space curves. Application of the theory of space curves to the power series expansions of the asymptotic curves leads to some interesting results. In the old coördinate system the equations of the osculating quadric cones at the point O of the u-curve and the v-curve are respectively

$$\begin{aligned} y^2 - (\tfrac{3}{2}\beta x - \tfrac{1}{2}\phi y - 9H^{(u)}z)z &= 0, \\ x^2 - (\tfrac{3}{2}\gamma y - \tfrac{1}{2}\psi x - 9H^{(v)}z)z &= 0, \end{aligned} \tag{36}$$

where $H^{(u)}$ is the same as the H defined in equations (23) and $H^{(v)}$ is easily written therefrom. The polar plane of the v-tangent, $x = z = 0$, with respect to the osculating quadric cone of the u-curve is the plane

$$y + \tfrac{1}{4}\phi z = 0. \tag{37}$$

Similarly, the polar plane of the u-tangent, $y = z = 0$, with respect to the osculating quadric cone of the v-curve is

$$x + \tfrac{1}{4}\psi z = 0. \tag{38}$$

Thus *the polar plane of the u-tangent with respect to the osculating quadric cone of the v-curve intersects the polar plane of the v-tangent with respect to the osculating quadric cone of the u-curve in the first edge of Green.*

Another quadric cone can be obtained from the two osculating quadric cones of the asymptotic curves. For this purpose let us consider any line l through the point O and lying in the tangent plane to the surface at the point O. This line has the equations

$$y - nx = z = 0 \qquad (n \neq 0). \tag{39}$$

The polar planes of the line l with respect to the cones (36) are respectively

$$y + 1/4n(n\phi - 3\beta)z = 0 \quad \text{and} \quad x + \tfrac{1}{4}(\psi - 3n\gamma)z = 0. \tag{40}$$

Eliminating n from these equations, we find

$$(x + \tfrac{1}{4}\psi z)(y + \tfrac{1}{4}\phi z) - \tfrac{9}{16}\beta\gamma z^2 = 0. \tag{41}$$

The following theorem summarizes this result:

Consider any line l through the point O of a surface and lying in the tangent plane of the surface at the point. Consider the polar planes of the line l with respect to the two osculating quadric cones of the asymptotic curves through the point O. As l varies about the point O the line of intersection of these two planes generates a quadric cone with its vertex at the point O.

It may be remarked that Calapso has obtained[1] the cone (41) by a method based on a consideration of the osculating twisted cubics of the asymptotic curves at the point 0 of the surface. He shows, among other things, that the polar line of the tangent plane of the surface at the point 0 with respect to this cone is the first edge of Green.

---

[1] R. Calapso, "Sugli enti proiettivi legati al generico punto di una superficie," *Atti Accademia Gioenia*, Catania, (5), vol. 19 (1932), mem. 14.

From the canonical expansions (29) we find that the u-curve belongs to a linear complex in case $B_6 = 0$, and is a twisted cubic in case $B_6 = A_6 = 0$. In the new coördinate system the principal plane at the point O of the u-curve is found to have the equation

$$B_6 Y - A_6 Z = 0, \tag{42}$$

and the principal point on the u-tangent has the coördinates

$$(3A_6,\ B_6,\ 0,\ 0). \tag{43}$$

If the u-curve belongs to a linear complex it is obvious that the principal plane coincides with the osculating plane of the curve and that the principal point coincides with the point O. Moreover, the principal plane is indeterminate in case the u-curve is a twisted cubic.

The Halphen cone at the point O of the u-curve is the eleven-point cubic cone with its vertex at the point O and with the u-tangent for a double line. In the new coördinate system *the equation of the Halphen cone at the point O of the u-curve is found to be*

$$X\left(YZ - \frac{B_7 - 2A_6}{B_6} Z^2\right) - Y^3 + \left(\frac{B_7 - 2A_6}{B_6}\right) Y^2 Z - B_6 Z^3 = 0. \tag{44}$$

The nodal plane of this cone has the equation

$$B_6 Y - (B_7 - 2A_6) Z = 0. \tag{45}$$

This plane coincides with the principal plane defined by equation (42) if, and only if,

$$B_7 - 3A_6 = 0. \tag{46}$$

Referring to equation (34), we thus obtain the following geometrical interpretation of the vanishing of the invariant θ_4:

The principal plane at a point of the u-curve coincides with the nodal plane of the Halphen cone at the point if, and only if, $\theta_4 = 0$.

Another geometrical interpretation for the condition $\theta_4 = 0$ can be given in the following way. The plane containing the three inflexional generators of the Halphen cone is found to have the equation

$$B_6^2 X - 2B_6(B_7 - 2A_6) Y + (B_7 - 2A_6)^2 Z = 0. \tag{47}$$

The pencil of quadric surfaces containing the osculating twisted cubic and having seven-point contact with the u-curve at the point 0 is given by

$$(48) \qquad h[Y - X^2 - (A_6/B_6)(Z - XY)] + k(Y^2 - ZX) = 0,$$

where h, k are arbitrary constants. Among these quadrics there are only two cones, for which

$$(49) \qquad h^2(kB_6^2 + hA_6^2)^2 = 0.$$

The one of these for which $h = 0$ is the osculating quadric cone

$$(50) \qquad Y^2 - ZX = 0$$

with its vertex at the point O. The other, for which

$$kB_6^2 + hA_6^2 = 0,$$

has the equation

$$(51) \qquad B_6^2(Y - X^2) - A_6B_6(Z - XY) - A_6^2(Y^2 - ZX) = 0,$$

and its vertex is at the Halphen point

$$(52) \qquad (A_6^3,\ A_6^2B_6,\ A_6B_6^2,\ B_6^3).$$

The polar plane of the principal point (43) with respect to the cone (51) is found to have the equation

$$(53) \qquad B_6^2X - 2A_6B_6Y + A_6^2Z = 0.$$

Comparison of this equation and equation (47) shows that these two planes coincide in case $B_7 - 3A_6 = 0$. Thus the following theorem is proved:

At a point O of the u-curve the plane containing the three inflexional generators of the Halphen cone coincides with the polar plane of the principal point with respect to the cone defined by equation (51) *if, and only if,* $\theta_4 = 0$.

Finally, it is easy to establish the following theorem:

If the u-curve belongs to a linear complex, the Halphen cone is composite, one component being the osculating plane and the other being the osculating quadric cone. Moreover, the osculating quadric cone has eight-point contact with the curve, provided $B_7 - 2A_6 \neq 0$.

THE UNIVERSITY OF CHICAGO

SYSTEMS OF QUADRICS ASSOCIATED WITH A POINT OF A SURFACE.*

By Louis Green.

1. Introduction. One of the popular methods of studying the local geometry of an analytic surface immersed in a three-dimensional projective space is to examine the properties of algebraic osculants. For, these osculants can be used to obtain simple characterizations of many elements intimately related to the given surface. The osculating quadrics at a point of a surface are of particular importance in this respect, and it is the purpose of this paper to make a more comprehensive study of certain of these quadrics, obtaining as a result significant properties of the surface.

Among other things, the problem of determining the envelope of a two-parameter family of Moutard quadrics is solved, the singularity of the curve of intersection of the surface and one of its Moutard quadrics is investigated, certain individual members of each Moutard pencil of quadrics[1] are characterized, and three new pencils of quadrics which we call the Darboux-Segre pencils of quadrics are defined and discussed.

2. Analytic basis. The local coördinate tetrahedron of G. M. Green (with a modified unit point) is taken as the projective coördinate system throughout this paper, and the equations of the various configurations to be used are obtained in this section by transformation from Fubini's coördinate system.

Fubini's canonical differential equations of an analytic curved surface S are given by

$$x_{uu} = px + \theta_u x_u + \beta x_v, \qquad x_{vv} = qx + \gamma x_u + \theta_v x_v, \qquad (\theta = \log \beta\gamma).$$

Let 0 be an ordinary point of S, chosen as the vertex $(1, 0, 0, 0)$ of two local tetrahedrons, one of Fubini and the other of Green. If a point of the ambient three-space has the two sets of coördinates $(x_1, \cdots, x_4)$, $(X_1, \cdots, X_4)$ relative to these respective tetrahedrons, then the following relations exist:[2]

---

* Presented to the Society, April 9, 1937. Received by the Editors October 30, 1937.

[1] So called because each of these pencils contains a quadric of Moutard. These pencils were first studied by Su and Ichida, *Japanese Journal of Mathematics*, vol. 10 (1933), pp. 209-216.

[2] Lane, "Power series expansions in the neighborhood of a point on a surface," *Proceedings of the National Academy of Sciences*, vol. 13 (1927), pp. 808-813.

$$
(1)\quad \begin{aligned}
x_1 &= 4X_1 + 3l\phi X_2 + 3m\psi X_3 + (9/4)n(\phi\psi - 8\theta_{uv})X_4,\\
x_2 &= -12lX_2 - 9n\psi X_4,\\
x_3 &= -12mX_3 - 9n\phi X_4,\\
x_4 &= 36nX_4,
\end{aligned}
$$

where l, m, n, ϕ, ψ, have the values

$$
l = 1/(\beta^2\gamma)^{1/3}, \qquad m = 1/(\beta\gamma^2)^{1/3}, \qquad n = lm,
$$
$$
\phi = (\log \beta\gamma^2)_u, \qquad \psi = (\log \beta^2\gamma)_v.
$$

With the introduction of non-homogeneous coördinates X, Y, Z defined in the customary way, 0 becomes the origin of a local non-homogeneous coördinate system, and the coördinates of a point on S sufficiently "near" 0 satisfy a convergent power series expansion of the form

$$
(2)\quad \begin{aligned}
Z = XY + X^3 + Y^3 + AX^3Y + BXY^3 + C_0X^5 + C_1X^4Y&\\
+ C_2X^3Y^2 + C_3X^2Y^3 + C_4XY^4 + C_5Y^5 + \cdots&,
\end{aligned}
$$

where all the coefficients are absolute invariants of the surface, and in particular,

$$
(3)\quad \begin{aligned}
A &= -3m\psi/2, \qquad B = -3l\phi/2,\\
20C_0 &= 36l^2p - 9l^2\phi_u - 27m\psi - (9/4)l^2\phi^2 + 9l^2\phi\theta_u,\\
16C_1 &= 24n\psi_u - 12n\phi_v + 135n\phi\psi - 135,\\
16C_2 &= -135m^2q + 72m^2\psi_v + 27m^2\psi^2 - 180l\phi + 72m^2\psi(\beta_v/\beta),\\
16C_3 &= -135l^2p + 72l^2\phi_u + 27l^2\phi^2 - 180m\psi + 72l^2\phi(\gamma_u/\gamma),\\
16C_4 &= 24n\phi_v - 12n\psi_u + 135n\phi\psi - 135,\\
20C_5 &= 36m^2q - 9m^2\psi_v - 27l\phi - (9/4)m^2\psi^2 + 9m^2\psi\theta_v.
\end{aligned}
$$

For use in § 7 the formulas for the derivatives of the local coördinates $X_1, \cdots, X_4$ will now be obtained. Equations (1) are first solved for the X_i, the resulting equations are then differentiated with respect to $u(v)$, and the x_{iu} (x_{iv}) which appear are replaced by their known values [3] in terms of the x_i. The right members are then expressed in terms of the X_i by means of (1), and when simplified with the use of (3) take the form

$$
(4)\quad \begin{aligned}
36lX_{1u} &= 6BX_1 + (60C_0 - 72A)X_2 + (12C_4 - 45AB + 108)X_3\\
&\quad + (36C_2 - 108B - 54A^2)X_4,\\
36lX_{2u} &= 12X_1 + 2BX_2 + (12C_4 - 45AB + 216)X_4,\\
36lX_{3u} &= -36X_2 - 2BX_3 + (60C_0 - 108A)X_4,\\
36lX_{4u} &= 12X_3 - 6BX_4;\\
36mX_{1v} &= 6AX_1 + (12C_1 - 45AB + 108)X_2 + (60C_5 - 72B)X_3\\
&\quad + (36C_3 - 108A - 54B^2)X_4,\\
36mX_{2v} &= -2AX_2 - 36X_3 + (60C_5 - 108B)X_4,\\
36mX_{3v} &= 12X_1 + 2AX_3 + (12C_1 - 45AB + 216)X_4,\\
36mX_{4v} &= 12X_2 - 6AX_4.
\end{aligned}
$$

---

[3] Lane, *Projective Differential Geometry of Curves and Surfaces*, Chicago, 1932, p. 111.

By means of transformation (1) all the remaining results of this section are readily established.

The tangents of Darboux through 0 are given by

$$X_2^3 + X_3^3 = 0 = X_4;$$

the equation of a quadric of Darboux,

$$x_2 x_3 - x_1 x_4 + h x_4^2 = 0 \qquad (h = h(u, v)),$$

becomes

(5) $\qquad X_2 X_3 - X_1 X_4 + K X_4^2 = 0, \qquad K = (9/2\beta\gamma)(\theta_{uv} + 2h);$

the equations of a canonical line of the second kind,

$$x_1 - k\phi x_2 - k\psi x_3 = 0 = x_4,$$

have the form

$$X_1 - (B/2)(1 + 4k)X_2 - (A/2)(1 + 4k)X_3 = 0 = X_4,$$

and its polar line with respect to a quadric of Darboux joins 0 to the point

$$(0, (A/2)(1 + 4k), (B/2)(1 + 4k), 1).$$

If the corresponding equations of a plane in the above-mentioned coördinate systems are

$$\sum_1^4 \xi_i x_i = 0, \qquad \sum_1^4 \eta_i X_i = 0,$$

then the transformation of Čech in Fubini's system,

$$\rho x_1 = \xi_2 \xi_3 \xi_4 - j(\gamma \xi_2^3 + \beta \xi_3^3), \qquad \rho x_2 = -\xi_2 \xi_3^2,$$
$$\rho x_3 = -\xi_2^2 \xi_3, \qquad x_4 = 0 \qquad\qquad (j = \text{const.}),$$

is expressed in Green's system by the equations

$$\sigma X_1 = \eta_2 \eta_3 \eta_4 + 3j(\eta_2^3 + \eta_3^3), \qquad \sigma X_2 = -\eta_2 \eta_3^2,$$
$$\sigma X_3 = -\eta_2^2 \eta_3, \qquad X_4 = 0.$$

A curve on S which passes through 0 in a non-asymptotic direction and satisfies the differential equation

$$dv - \lambda du = 0 \qquad (\lambda = \lambda(u, v)),$$

has for its osculating plane at 0 the plane represented by

$$2\lambda(\lambda x_2 - x_3) + (\beta - \gamma\lambda^3 - \theta_u \lambda + \theta_v \lambda^2 + \lambda')x_4 = 0 \qquad (\lambda' = v'' = \lambda_u + \lambda\lambda_v).$$

Under transformation of coördinates (1) this equation becomes

$$(6) \qquad 6\mu(\mu X_2 - X_3) + (9\mu^3 - 9 + A\mu^2 - B\mu - 9\mu')X_4 = 0,$$

where μ, μ', are defined by

$$(7) \qquad \mu = \lambda l/m, \qquad \mu' = l(\mu_u + \lambda\mu_v) = l\mu_u + m\mu\mu_v.$$

The osculating plane at 0 of the pangeodesic which passes through this point in a non-Darboux direction $X_3/X_2 = \mu$ is expressed by

$$(8) \qquad 2\mu(\mu^3 + 1)(\mu X_2 - X_3) + (3\mu^6 - 3 + A\mu^2 - B\mu^4)X_4 = 0,$$

and the quadric of Moutard in the non-asymptotic direction μ at 0 has the equation

$$(9) \qquad \mu^3(X_2X_3 - X_1X_4) + \mu^2(2 - \mu^3)X_2X_4 + \mu(2\mu^3 - 1)X_3X_4 \\ + (A\mu^2 + B\mu^4 - \mu^6 - 2\mu^3 - 1)X_4^2 = 0.$$

3. The curve of intersection of a surface and a quadric. The equation of the most general quadric having second-order contact with the surface S at the point 0 is

$$(10) \qquad XY - Z + K_2XZ + K_3YZ + K_4Z^2 = 0,$$

where K_2, K_3, K_4 are arbitrary functions of u and v. The curve of intersection C of this quadric and S is expressed by (10) and

$$(11) \qquad 0 = X^3 - K_2X^2Y - K_3XY^2 + Y^3 + (A - K_2^2)X^3Y \\ - (2K_2K_3 + K_4)X^2Y^2 + (B - K_3^2)XY^3 + \cdots,$$

the latter equation being obtained by solving (10) for Z as a power series in X, Y and then subtracting from (2). The curve C has a triple-point at 0 with triple-point tangents satisfying the equations

$$(12) \qquad X^3 - K_2X^2Y - K_3XY^2 + Y^3 = 0 = Z.$$

Comparison of equations (10) and (12) leads to the following conclusion:

Two arbitrary non-asymptotic tangents to S at 0 determine a pencil of quadrics having second-order contact with S at 0 and meeting S in a one-parameter family of curves each containing these tangents as two of its triple-point tangents. The residual tangent at 0 is the same for all the curves of the family.

A particular quadric of such a pencil is uniquely determined, ordinarily, by characterizing the osculating plane at 0 of *any one* of the three branches of the curve C.[4]

---

[4] The Moutard pencils of quadrics are the only pencils for which this is not the case. But, as is shown in §§ 5, 6, for such a pencil of quadrics in a direction μ the

Several particular examples of the above theorem are worth noting. If two of the triple-point tangents lie in distinct Darboux directions at 0, the third tangent must lie in the residual Darboux direction and the resulting quadrics are the quadrics of Darboux.

If two of the tangents coincide in a non-asymptotic direction $X_3/X_2 = \mu$, the third must lie in the direction $-1/\mu^2$, and the quadrics determined form the Moutard pencil of quadrics in the direction μ. If $\mu^3 = -1$ the triple-point tangents all coincide in one of the Darboux directions, while if $\mu^3 = 1$ two of the tangents coincide in a Segre direction and the third tangent lies in the conjugate Darboux direction.

If two of the tangents lie in distinct Segre directions, the third is seen to lie in that Darboux direction which is conjugate to the residual Segre direction. Every Darboux direction thus determines a different pencil of such quadrics, so that each of the three pencils may be designated as the Darboux-Segre pencil of quadrics in the particular Darboux direction considered.

The Darboux quadrics and the three Darboux-Segre pencils of quadrics are special cases of the pencil

$$XY - Z - \frac{\mu^3+1}{\mu} XZ - \frac{\mu^3+1}{\mu^2} YZ + K_4 Z^2 = 0 \qquad (K_4 \text{ arbitrary}),$$

which is determined by the tangents in the directions $\mu\omega, \mu\omega^2, -1/\mu^2$, where ω is a primitive cube root of unity.

The quadrics of Darboux are thus seen to constitute the only pencil which can be determined by three tangents apolar to the asymptotic tangents at 0. Similarly, there exists a unique triple of tangents apolar to a given pair of conjugate tangents $X_3/X_2 = \pm \mu$, $X_4 = 0$, which according to the above theorem can determine a pencil of quadrics. These quadrics, whose equations can be shown to have the form

$$XY - Z - 3\mu^2 XZ - (3/\mu^2) YZ + K_4 Z^2 = 0, \qquad (K_4 \text{ arbitrary}),$$

were defined by Davis [5] in an entirely different manner.

4. Quadrics associated with the tangents of Darboux and of Segre. The theorem of the last section is applied here to the quadrics of Darboux, the Moutard pencils in the Segre directions, and the Darboux-Segre pencils of quadrics.

---

osculating plane at 0 of a properly chosen branch of C does determine a unique quadric of the pencil, except when μ is a Darboux direction. In this latter case the osculating plane at 0 of each of the three branches of C coincides with the tangent plane to S, unless the quadric is the quadric of Moutard of the pencil.

[5] "Contributions to the theory of conjugate nets," *Chicago Dissertation* (1933), p. 12.

The surface S and an arbitrary quadric of Darboux with parameter K meet in a curve C whose three branches are representable, as is seen from (2) and (5), by the series

$$Y = -\epsilon_i X + \tfrac{1}{3}(A\epsilon_i^2 + B\epsilon_i + K)X^2 + mX^3 + \cdots,$$
$$Z = -\epsilon_i X^2 + \tfrac{1}{3}(A\epsilon_i^2 + B\epsilon_i + K)X^3 + \cdots, \qquad (i = 1, 2, 3),$$

where

$$-9\epsilon_i^2 m = 2AB\epsilon_i + 2B^2 + AK + 3BK\epsilon_i^2 + K^2\epsilon_i + 3\sum_{j=0}^{5} C_j(-\epsilon_i)^j,$$
$$(13) \qquad \epsilon_1 = 1, \quad \epsilon_2 = \omega, \quad \epsilon_3 = \omega^2$$

and ω is a primitive cube root of unity. The osculating plane at 0 of each branch of C has the equation

$$(14) \qquad \epsilon_i^2 X + \epsilon_i Y + \tfrac{1}{3}(A\epsilon_i^2 + B\epsilon_i + K)Z = 0,$$

from which it follows that *these osculating planes are coaxial if, and only if, the given quadric is the quadric of Wilczynski* ($K = 0$). *The common axis of the three planes then becomes the canonical line of the first kind for which* $k = -5/12$, *which we shall designate as the first axis of Bompiani.*[6]

It may be remarked here that the Moutard quadric in a Darboux direction $-\epsilon_i$ meets a Darboux quadric (5) in the asymptotic tangents through 0 and in a conic lying in the plane (14). As a consequence the Moutard quadrics in the three Darboux directions have a residual point of intersection on the first axis of Bompiani.

If the parameter K is left arbitrary, then the planes (14) correspond under a transformation of Čech with parameter j to three collinear points if, and only if, $K = -18j$ (the line thereby determined being the polar with respect to a Darboux quadric of the first axis of Bompiani, independently of the value of K). This equation defines a correspondence at each point of the surface between the Čech transformations and those Darboux quadrics for which K is independent of u, v. Thus, the quadrics of Lie and Fubini for which $K = -9/2$ and -3 respectively, characterize the transformations of Čech for which $j = 1/4$ and $1/6$, whereas the transformation of Segre ($j = 1$) characterizes the quadric of Darboux for which $K = -18$.

The osculating plane (14) is a stationary osculating plane if, and only if,

$$-9\epsilon_i m = (A\epsilon_i^2 + B\epsilon_i + K)^2;$$

the value of K is thereby uniquely determined[7] and when substituted into

---

[6] This line was first characterized in a more complicated way by Bompiani in the *Rendiconti dei Lincei*, ser. 5, vol. 33_1 (1924), pp. 85-90.

[7] The value of K depends on the particular Darboux direction chosen, so that there exist three distinct Darboux quadrics with the property considered.

(14) gives as the equation of the stationary osculating plane

$$(15) \qquad (A\epsilon_i - B)(\epsilon_i X + Y) + [\sum_{j=0}^{5} C_j(-\epsilon_i)^j]Z = 0.$$

The Moutard pencil of quadrics at 0 in the Segre direction $X_3/X_2 = \epsilon_i$ has the form

$$(16) \qquad XY - Z + \epsilon_i^2 XZ + \epsilon_i YZ + K_i Z^2 = 0,$$

with K_i as parameter of the pencil. The curve of intersection of the surface S and an arbitrary quadric of this pencil has a branch in the conjugate Darboux direction, the osculating plane at 0 of this branch being expressed by

$$(17) \qquad \epsilon_i^2 X + \epsilon_i Y + \tfrac{1}{4}(A\epsilon_i^2 + B\epsilon + K_i)Z = 0.$$

Since the equation of the osculating plane at 0 of the curve of Darboux in the direction $-\epsilon_i$ is

$$(18) \qquad \epsilon_i^2 X + \epsilon_i Y - \tfrac{1}{6}(A\epsilon_i^2 + B\epsilon_i - 18)Z = 0,$$

it follows that the planes (17) and (18) coincide only if

$$K_i = -(A/3)\epsilon_i^2 - (B/3)\epsilon_i - 12.$$

This value of K_i determines a unique quadric of the pencil (16), and since there are three Moutard pencils in the Segre directions at 0, three corresponding quadrics are obtained in all. *These quadrics intersect in the asymptotic tangents through* 0 *and in a residual point lying on the canonical line of the first kind for which* $k = -1/12$. *Under the transformation of Čech with* $j = 1/2$ *the three planes* (18) *are carried into collinear points on the second axis of Čech.*

The condition that the osculating plane (17) contain a line l joining 0 to an arbitrary point $(0, a, b, 1)$ requires that

$$K_i = -(4a + A)\epsilon_i^2 - (4b + B)\epsilon_i.$$

The line l obtains in this way a single quadric out of each of the three Moutard pencils in the Segre directions, the residual point of intersection P of these quadrics having the coördinates

$$(4a + A)(4b + B), \quad 4a + A, \quad 4b + B, \quad 1.$$

As the line l *varies in the bundle at* 0, *the point* P *generates the quadric of Wilczynski;* P *lies on* l *if, and only if,* l *is the first axis of Bompiani.*

The quadric of Moutard in the Segre direction ϵ_i at 0 is that quadric of the pencil (16) for which

$$K_i = A\epsilon_i^2 + B\epsilon_i - 4;$$

the osculating plane (17) determined by this value of K_i can also be described as the plane containing the non-degenerate conic of intersection of the Moutard quadrics in the conjugate directions $\pm\epsilon_i$. The residual point of intersection of the Moutard quadrics in the three Segre directions lies on the canonical line of the first kind for which $k = -3/4$, while the transformation of Čech for which $j = 1/6$ carries the three planes (17) determined by these quadrics into collinear points on the second directrix of Wilczynski.

The Darboux-Segre pencils of quadrics will now be considered. For each choice of ϵ_i let $\epsilon_i, \epsilon_j, \epsilon_k$ represent the three values in (13). Then the three directions

$$X_3/X_2 = -\epsilon_i, \qquad X_3/X_2 = \epsilon_j, \qquad X_3/X_2 = \epsilon_k$$

at 0 determine a Darboux-Segre pencil of quadrics in the Darboux direction $-\epsilon_i$, having the equation

$$(19) \quad XY - Z - 2\epsilon_i^2 XZ - 2\epsilon_i YZ + K_i Z^2 = 0 \qquad (K_i \text{ arbitrary}).$$

If the osculating planes at 0 of the three branches of the curve of intersection of the surface S and one of these quadrics be denoted by $\pi_{ii}, \pi_{ij}, \pi_{ik}$ corresponding to the respective directions $-\epsilon_i, \epsilon_j, \epsilon_k$, then

$$\pi_{ii}: \quad \epsilon_i^2 X + \epsilon_i Y + (A\epsilon_i^2 + B\epsilon_i + K_i)Z = 0,$$

$$\pi_{ih}: \quad \epsilon_h^2 X - \epsilon_h Y + P_{ih} Z = 0 \qquad (h = j, k),$$

where

$$P_{ih} = \frac{-A\epsilon_h - B + 4\epsilon_h^2 + K_i\epsilon_h^2}{2\epsilon_i^2 + 4\epsilon_i\epsilon_h + 3\epsilon_h^2}.$$

The condition that the plane π_{ii} coincide with the osculating plane of a curve of Darboux at 0 determines a unique quadric out of each Darboux-Segre pencil. *The residual point of intersection of these three quadrics is the point of intersection (besides* 0) *of the quadric of Fubini with the canonical line of the first kind for which* $k = -11/24$.

A unique quadric in each Darboux-Segre pencil is also obtained from the condition that the three planes $\pi_{ii}, \pi_{ij}, \pi_{ik}$ be coaxial, the parameters K_i then having the values

$$K_i = -\tfrac{4}{5}(A\epsilon_i^2 + B\epsilon_i + 2) \qquad (i = 1, 2, 3).$$

The three quadrics thus found have their residual point of intersection on the canonical line of the first kind for which $k = -9/20$.

The locus of the line of intersections of the planes π_{ij}, π_{ik} for all quadrics in the pencil (19) is a plane passing through the Segre tangent $X_3/X_2 = \epsilon_i$, $X_4 = 0$ and through the canonical line of the first kind for which $k = -3/4$. The quadric determined by the condition that the planes π_{ij}, π_{ik} contain this canonical line corresponds to the parametric value

$$K_i = -2A\epsilon_i^2 - 2B\epsilon_i - 4,$$

and the three quadrics so obtained for $i = 1, 2, 3$ have their residual point of intersection on this same line.

5. The Moutard pencils of quadrics. This section is concerned essentially with an examination of the singularity of the curve of intersection of the surface S and a quadric of the Moutard pencil. The curve determined by the Moutard quadric of the pencil is found to differ radically from the curves determined by the other quadrics of the pencil.

The Moutard pencil of quadrics determined at a point 0 of S by a direction $X_3/X_2 = \mu$ has the form

$$(20) \qquad XY - Z + (1/\mu)(2 - \mu^3)XZ + (1/\mu^2)(2\mu^3 - 1)YZ + K_\mu Z^2 = 0,$$

where K_μ is an arbitrary function of u, v. The quadric of Moutard of the pencil is obtained when

$$(21) \qquad \mu^3 K_\mu = A\mu^2 + B\mu^4 - (\mu^3 + 1)^2.$$

Let the equation

$$(22) \qquad Z = n(Y - \mu X) \qquad (n \neq 0)$$

represent an arbitrary plane through the tangent $X_3/X_2 = \mu$, $X_4 = 0$, cutting the surface S in a curve s and a quadric of the pencil (20) in a conic q. Then s and q have third-order contact at 0 for all values of n and K_μ. Conversely, if every plane of the form (22) meets the surface S and a quadric Q in a pair of curves which have contact of the third order at 0, then Q belongs to the pencil (20). The quadric of Moutard (9) is the unique quadric of the pencil for which the contact between s and q is of the fourth order for all values of n.

If μ is not a Darboux direction, there are two planes of sextactic section[8] passing through the tangent $X_3/X_2 = \mu$, $X_4 = 0$, for which the contact between s and q is of the fifth order at 0. The values of n in (22) which determine these planes are the roots of the equation

$$(23) \qquad hn^2 + (3 - 3\mu^6 - A\mu^2 + B\mu^4)\mu n + (\mu^3 + 1)\mu^2 = 0,$$

---

[8] Darboux, *Bulletin des Sciences Mathematiques et Astronomiques*, ser. 2, vol. 4 (1880), p. 356.

where h is defined by

$$h = 2(\mu^3 + 1)^3 - 3\mu^2(\mu^3 + 1)(A + B\mu^2) + \mu^2\alpha,$$

and

$$\alpha = \sum_0^5 C_i\mu^i. \tag{24}$$

The curve of intersection of S and an arbitrary quadric of the pencil (20) is given by equations (20) and (11). If μ is not a Darboux direction, this curve has two branches at 0 in the direction μ, whose equations are

$$Y = \mu X \pm rX^{3/2} + \cdots, \qquad Z = XY + \cdots, \tag{25}$$

where r satisfies the equation

$$r^2 = -\mu(A\mu^2 + B\mu^4 - \mu^6 - 2\mu^3 - 1 - \mu^3 K_\mu)/(\mu^3 + 1).$$

If $r = 0$ the curve (25) is found to have a tacnode in the direction μ, so that in view of (21) *a necessary and sufficient condition that a quadric of the Moutard pencil in a non-Darboux direction* μ *at* 0 *be the quadric of Moutard* (9) *is that its curve of intersection with* S *have a tacnode at* 0 *in the direction* μ *instead of a cusp.*

For the case of the Moutard quadric (9) the terms of fifth degree in the series (11) can be shown to be

$$\sum_0^5 T_i X^{5-i} Y^i$$

where

$$\begin{aligned}
&T_0 = C_0, \qquad \mu^3 T_1 = \mu^3 C_1 + \mu^9 - 6\mu^6 + 12\mu^3 - 8,\\
&\mu^4 T_2 = \mu^4 C_2 - 9\mu^9 + 3B\mu^7 + 27\mu^6 + 3A\mu^5 - 6B\mu^4 - 27\mu^3 - 6A\mu^2 + 18,\\
&\mu^5 T_3 = \mu^5 C_3 + 18\mu^9 - 6B\mu^7 - 27\mu^6 - 6A\mu^5 + 3B\mu^4 + 27\mu^3 + 3A\mu^2 - 9,\\
&\mu^6 T_4 = \mu^6 C_4 - 8\mu^9 + 12\mu^6 - 6\mu^3 + 1, \qquad T_5 = C_5.
\end{aligned}$$

The two tacnode branches of the curve of intersection of the surface S and the Moutard quadric (9) are now found to have the equations

$$Y = \mu X + \sigma_i X^2 + \cdots, \qquad Z = XY + \cdots \qquad (i = 1, 2), \tag{26}$$

where σ_1, σ_2 are the roots of the equation

$$(\mu^3 + 1)\sigma^2 + (3 - 3\mu^6 - A\mu^2 + B\mu^4)\sigma + h = 0. \tag{27}$$

If $r \neq 0$, the osculating plane at 0 of each cusp-branch (25) coincides with the tangent plane $Z = 0$, whereas if $r = 0$ the osculating planes at 0 are represented by the equations

$$Z = (\mu/\sigma_i)(Y - \mu X) \qquad (i = 1, 2). \tag{28}$$

Setting $\mu/\sigma_i = n_i$ and using (27) we find that n_1, n_2 are the roots of equation (23). Hence, *the osculating planes at* 0 *of the tacnode branches of the curve of intersection of the surface S and the Moutard quadric in the non-Darboux direction μ at* 0 *coincide with the planes of sextactic section which pass through the tangent line in the direction μ.*

Comparison of the discriminants of (23) and (27) shows that the two planes of sextactic section through the line $X_3/X_2 = \mu$, $X_4 = 0$ coincide if, and only if, the tacnode branches (26) have maximum-order contact at 0, this situation occurring for twelve values[9] of μ.

The results concerning the Moutard pencil of quadrics in a Darboux direction $X_3/X_2 = -\epsilon_i$ at 0 will be stated briefly. Each quadric Q of the pencil meets the surface S in a curve C whose triple-point tangents all coincide in the given Darboux direction. If Q is not the Moutard quadric of the pencil, this curve has a superlinear branch of order 3 and the osculating plane at 0 of each branch coincides with the plane $Z = 0$. If Q is the Moutard quadric then the curve C consists of two cusped branches and one linear branch. The osculating planes of the cusped branches at 0 coincide with the tangent plane to S, while the osculating plane of the linear branch at 0 is none other than the plane (15). This plane can also be obtained as a limiting case of (28) as μ approaches $-\epsilon_i$.

6. Cones associated with the Moutard pencils of quadrics. The problem of characterizing individual members of the Moutard pencils of quadrics in the Segre directions at a point of a surface has already been discussed. We now consider this problem for an arbitrary direction.

The curve of intersection of the surface S and an arbitrary quadric of the Moutard pencil (20) in a non-Darboux direction μ at 0 has a linear branch in the direction $-1/\mu^2$. The equations of this branch are

$$Y = -\frac{1}{\mu^2}X + \frac{A\mu^8 + B\mu^4 - (\mu^6 - 1)^2 + \mu^6 K_\mu}{\mu^6(\mu^3 + 1)^2}X^2 + \cdots, \qquad Z = XY + \cdots,$$

so that its osculating plane at 0 is of the form

$$\mu^2(\mu^3 + 1)(X + \mu^2 Y) + [A\mu^8 + B\mu^4 - (\mu^6 - 1)^2 + \mu^6 K_\mu]Z = 0. \tag{29}$$

The condition that this plane contain a line l joining 0 to an arbitrary point $(0, a, b, 1)$ determines a unique quadric Q_μ of the pencil, since the parameter K_μ must satisfy the equation

---

[9] Wilczynski, "Fifth memoir," *Transactions of the American Mathematical Society*, vol. 10 (1909), p. 293.

$$(30)\qquad \mu^6 K_\mu = (\mu^6 - 1)^2 - A\mu^8 - a(\mu^3 + 1)^2\mu^2 - B\mu^4 - b(\mu^3 + 1)^2\mu^4.$$

Associated with the direction μ are two others, $\mu\omega$, $\mu\omega^2$ which form with μ a triple of directions apolar to the asymptotic directions. Thus two other quadrics, $Q_{\mu\omega}$, $Q_{\mu\omega^3}$, belonging to the Moutard pencils at 0 in the directions $\mu\omega$, $\mu\omega^2$, can be defined by this line l in exactly the same way as is Q_μ, the equations of these quadrics being obtained from (20) and (30) by replacing μ by $\mu\omega$, $\mu\omega^2$, respectively. The three quadrics Q_μ, $Q_{\mu\omega}$, $Q_{\mu\omega^2}$ intersect in the asymptotic tangents through 0 and in a residual point whose join with 0 is a line l' containing the point

$$(31)\qquad \begin{aligned} &X_1 = 0, \qquad X_2 = (2\mu^3 - 1)[A\mu^6 + a(\mu^3 + 1)^2], \\ &X_3 = \mu^3(2 - \mu^3)[B + b(\mu^3 + 1)^2], \qquad X_4 = \mu^3(2 - \mu^3)(2\mu^3 - 1). \end{aligned}$$

The lines l and l' coincide if, and only if,

$$(32)\qquad a = -A\mu^6/(2\mu^6 + 1), \qquad b = -B/(\mu^6 + 2),$$

so that the line l and the quadric Q_μ are uniquely determined by the given direction μ. As μ varies, l generates the quadric cone given by

$$(33)\qquad 3XY + 2AYZ + 2BXZ + ABZ^2 = 0.$$

If, however, the line l is kept fixed and the direction μ is varied, then l' generates a cone which in general is of the fourth order but for special positions of l may be of the second order. To determine the conditions under which this occurs, we note that any quadric cone generated by the line l' must be of the form

$$(34)\qquad XY - pYZ + qXZ + rZ^2 = 0,$$

where p, q, r are functions of u, v. Substitution of equations (31) into (34) leads to seven equations of which the following five are independent:

$$\begin{aligned} &b(p - a - A) = 0, \qquad a(q - b - B) = 0, \\ &Ab - 3ab - 5pb - 4q(a + A) + 4r = 0, \\ &aB - 3ab - 5aq - 4p(b + B) + 4r = 0, \\ &5AB + Ab + aB + 14ab + p(11b - 9B) + (11a - 9A)q + 33r = 0. \end{aligned}$$

These equations have five sets of solutions for a, b, p, q, r which are exhibited in the following table:

Case	a	b	p	q	r
(I)	$-A/3$	$-B/3$	$2A/3$	$2B/3$	$AB/3$
(II)	0	0	$-A/3$	$-B/3$	$-AB/3$
(III)	$-4A/9$	$-4B/9$	$5A/9$	$5B/9$	$7AB/27$
(IV)	0	$-4B/9$	A	B	$5AB/9$
(V)	$4A/9$	0	A	B	$5AB/9$

Hence, *a necessary and sufficient condition that the line l' generate a quadric cone as μ varies is that the fixed line l, which is the axis of the pencil of planes* (29), *join* 0 *to the point* $(0, a, b, 1)$ *where a, b have any one of the five pairs of values listed in the table.*

Five quadrics belonging to the Moutard pencil in an arbitrary non-Darboux direction[10] μ at 0 are thus characterized, each quadric having associated with it a quadric cone and a fixed line l. In cases (I) and (II) of the table this line is the first axis of Bompiani and the first edge of Green respectively, while in case (III) it is the canonical line of the first kind for which $k = -17/36$. The lines determined in cases (IV) and (V) can be described as the polars with respect to a quadric of Darboux of the lines A_XB_Y, A_YB_X, where A_X, A_Y and B_X, B_Y are the intersections of the asymptotic tangents through 0 with the canonical lines of the second kind for which $k = -17/36, -1/4$.

The quadric cones associated with these five quadrics are all mutually tangent along the first axis of Bompiani, the common tangent plane passing through the second canonical tangent. Consideration of these cones readily leads to the characterization of a number of new canonical lines of the first kind.

The cone (33), which is also one of the cones listed in the table, can be obtained in still another way. The line joining 0 to the residual point of intersection of the three quadrics of Moutard in the directions μ, $\mu\omega$, $\mu\omega^2$ generates the cone (33) as μ varies.[11]

There is a certain cone of class 3 closely associated with the Moutard quadrics at 0. The quadrics of Moutard in the conjugate directions μ, $-\mu$ meet in the asymptotic tangents through 0 and in a conic whose plane has the equation

$$2\mu^2(X + \mu^2 Y) + (A\mu^2 + B\mu^4 - \mu^6 - 1)Z = 0.$$

As μ varies this plane envelopes the cone, which in local plane coördinates $\eta_1, \cdots, \eta_4$ has the equations

---

[10] If μ is a Darboux direction these quadrics, as well as the quadric determined by (32), all coincide with the Moutard quadric.

[11] Bompiani, "Contributi alla geometria proiettivo-differenziale di una superficie," *Bollettino della Matematica Italiana*, vol. 3 (1924), pp. 97-100.

$$\eta_2^3+\eta_3^3-\eta_2\eta_3(A\eta_2+B\eta_3-2\eta_4)=0=\eta_1.$$

The cusp-planes of this cone pass through the Segre tangents at 0 and intersect in the first directrix of Wilczynski.

7. Envelopes of Moutard quadrics. The problem of determining the envelope of a family of Moutard quadrics is of sufficient importance to warrant consideration. This problem has been discussed only by Kimpara,[12] to the best of the writer's knowledge, and then merely for a special case.

The one-parameter family of Moutard quadrics at a point 0 of a surface S has an enveloping surface Σ which contains the asymptotic tangents through 0 as a locus of singular points. Elimination of μ between (9) and its μ-derivative shows that Σ is an algebraic surface of order 14. The characteristic curve determined by Σ and a particular quadric of the family in the direction μ is a conic lying in the osculating plane at 0 of the pangeodesic which passes through 0 in the direction μ. Hence, *the lines joining* 0 *to the points on the edge of regression of* Σ *generate the cone of Segre.*

We now proceed with the problem of determining the envelope of a two-parameter family of Moutard quadrics. Let $C^{(1)}$ be a one-parameter family of curves on S which satisfy the differential equation

$$dv-\lambda du=0 \qquad (\lambda=\lambda(u,v)), \tag{35}$$

and let it be assumed that these curves simply cover the portion of the surface under consideration and are nowhere tangent to an asymptotic curve. The tangents to $C^{(1)}$ then determine a two-parameter family of Moutard quadrics $Q^{(2)}$, one at each point of S. A particular point 0 therefore has associated with it a curve C_μ of the family $C^{(1)}$, whose tangent at 0 is

$$X_3/X_2=\mu, \quad X_4=0 \qquad (\mu=\lambda l/m),$$

and a quadric Q_μ of the family $Q^{(2)}$. An arbitrary curve C_ν on S, which passes through 0 in a direction $X_3/X_2=\nu$, determines at each of its points a unique Moutard quadric belonging to $Q^{(2)}$, the quadrics so obtained forming a one parameter family $Q^{(1)}$ with an envelope E_ν. The quadric Q_μ at 0 is a member of this family and hence touches E_ν along a characteristic curve Γ_ν.

These curves Γ_ν and the envelope E of the family of quadrics $Q^{(2)}$ depend on the given curves $C^{(1)}$, and their properties differ widely in two separate cases, namely, when the curves $C^{(1)}$ are curves of Darboux and when they are nowhere tangent to a curve of Darboux.

---

[12] "Sur l'enveloppe des quadriques de Moutard," *Ryojun College of Engineering, Commemoration volume* (1934), pp. 33-39.

(i) Suppose first that the curves $C^{(1)}$ form a one-parameter family of curves of Darboux. The quadric Q_μ is the Moutard quadric in the Darboux direction $X_3/X_2 = \mu$ at 0, and has the equation

$$(36) \qquad G \equiv X_2X_3 - X_1X_4 - 2\mu^2X_2X_4 + 3\mu X_3X_4 + (B\mu - A\mu^2)X_4^2 = 0.$$

The characteristic curve Γ_ν determined by Q_μ and a curve C_ν through 0 is therefore given by (36) and

$$G_u + G_v(dv/du) = 0,$$

where dv/du refers to the direction of the tangent to C_ν at 0, and since $dv/du = \nu m/l$ this may be written

$$(37) \qquad lG_u + \nu mG_\nu = 0.$$

The derivatives of A and B which appear in this equation may be readily evaluated. For,

$$A = -(3/2)m\psi, \qquad A_u = -(3/2)(m_u\psi + m\psi_u),$$
$$m = 1/(\beta\gamma^2)^{1/3}, \qquad m_u = -(m/3)(\beta_u/\beta) - (2m/3)(\gamma_u/\gamma),$$

and consequently,

$$18lA_u = 4AB - 27m\psi_u.$$

Now from (3) it is seen that $m\psi_u$ can be expressed in terms of the coefficients appearing in series (2), and hence lA_u is expressible in like fashion. Thus we find

$$36lA_u = 273AB - 648 - 48C_1 - 24C_4,$$
$$36mB_v = 273AB - 648 - 48C_4 - 24C_1,$$
$$36mA_v = 52A^2 + 288B - 24C_2 - 120C_5,$$
$$36lB_u = 52B^2 + 288A - 24C_3 - 120C_0.$$

With the aid of formulas (4) the left member of (37) can now be expressed in terms of μ, ν, the coefficients of series (2), and the local coördinates $X_1, \cdots, X_4$. Elimination of the term X_1X_4 from the resulting equation by means of (36) finally yields the simplified result

$$(38) \qquad 36(\mu - \nu)(\mu X_2 - X_3) + LX_2X_4 + MX_3X_4 + NX_4^2 = 0,$$

where the coefficients L, M, N have the values

$$L = 12B\mu(\mu + 2\nu) - 36A + 216(\nu - \mu),$$
$$M\mu = 12\mu(A\mu + B)(\mu - \nu) - L,$$
$$N = 72A(2\nu - \mu) - 72B\mu^2(2\mu - \nu) + B^2(54\nu + 40\mu) - A^2\mu^2(54\mu + 40\nu)$$
$$+ 131AB\mu(\nu - \mu) + 12\sum_0^5 C_i\mu^i[(5 - i)\mu + i\nu].$$

If $\nu \neq \mu$, the equations of Γ_ν can also be written in the non-homogeneous form

$$Y = \mu X \pm \tfrac{1}{6}\left(\frac{L\mu + M\mu^2}{\nu - \mu}\right)^{\frac{1}{2}} X^{3/2} + \cdots, \qquad Z = XY + \cdots \qquad (\nu \neq \mu),$$

which simplify to

$$Y = \mu X \pm [\tfrac{1}{3}(A - B\mu^2)]^{\frac{1}{2}} X^{3/2} + \cdots, \qquad Z = XY + \cdots \qquad (\nu \neq \mu).$$

Hence, if $\nu \neq \mu$, the characteristic curve Γ_ν is a quartic curve with a cusp at 0 and a cusp-tangent in the Darboux direction μ.

If C_ν is tangent at 0 to the curve of Darboux C_μ, then the characteristic curve Γ_ν consists of the asymptotic tangents through 0 and of a non-degenerate conic lying in the plane whose equation is

$$\begin{aligned}(39) \qquad & 18(A\mu + B)(\mu X_2 - X_3) \\ & + [30\alpha - 36\mu^2(A\mu + B) - 47(A^2\mu^2 - B^2)]X_4 = 0,\end{aligned}$$

α being defined in (24). Since this conic is tangent to C_μ at 0, it follows that if C_ν and C_μ coincide, then C_μ is part of the edge of regression of the envelope E_ν.

The points where the Quadric Q_μ touches the envelope E of the two-parameter family of Moutard quadrics $Q^{(2)}$ are determined by the simultaneous solutions of equations (36), (38) and (39), where in (38) the value of ν is arbitrary but different from μ. Since the plane (39) intersects the cone (38) in the Darboux tangent $X_3/X_2 = \mu$, $X_4 = 0$ and in another line l which passes through 0, and since these two lines meet the quadric Q_μ at 0 and at some other point P not lying on S, it therefore follows that *the envelope E consists of the surface S and of another sheet generated by the point P.*

(ii) We now assume that the curves $C^{(1)}$ are nowhere tangent to a curve of Darboux. If $H = 0$ is the equation (9) of the Moutard quadric Q_μ in the direction μ at 0, then the characteristic curve Γ_ν determined by Q_μ and a curve C_ν through 0 has the form

$$H = 0, \qquad lH_u + \nu m H_v = 0.$$

The second of these equations can be expressed in terms of μ, ν, the coefficients of series (2), and the local coördinates $X_1, \cdots, X_4$, and when combined with (9) to eliminate the term X_1X_4 becomes

$$\begin{aligned}(40) \qquad & 12\mu^4(\mu X_2 - X_3)[(2\nu - 3\mu - \mu^3\nu)X_2 - (2\mu^3 - 3\mu^3\nu - 1)X_3] \\ & + DX_2X_4 + EX_3X_4 + FX_4^2 = 0,\end{aligned}$$

where D, E, F have the values

$$\begin{aligned}
D = {} & 4A\mu^5(2\mu^3\nu - 9\mu + 2\nu) + 4B\mu^5(\mu^3 + 6\mu^2\nu - 2) \\
& + 12\mu^3[\mu^6(\mu - 4\nu) - 2\mu^3(\mu - \nu) + 7\mu - 4\nu] \\
& - 72\mu^4(\mu^3 + 1)(l\mu_u + \nu m\mu_v), \\
E\mu = {} & 12\mu^3(\nu - \mu)(3\mu^6 - 3 + A\mu^2 - B\mu^4) - D, \\
F = {} & 12A\mu^4[(5\mu^3 + 11) + \mu^2\nu(\mu^3 + 13)] \\
& + 12B\mu^3[(13\mu^3 + 1) + \mu^2\nu(11\mu^3 + 5)] \\
& + A^2\mu^5(54\mu + 40\nu) + B^2\mu^6(40\mu + 54\nu) \\
& + AB\mu^4[\mu(45\mu^3 + 176) + \nu(176\mu^3 + 45)] \\
& + 12\mu[\mu(\mu^9 - 18\mu^6 - 21\mu^3 - 2) + \nu(-2\mu^9 - 21\mu^6 - 18\mu^3 + 1)] \\
& - 12\mu^3 \sum_0^5 C_i\mu^i[(5 - i)\mu + i\nu] \\
& + 36\mu^2(3 - 3\mu^6 - A\mu^2 - B\mu^4)(l\mu_u + \nu m\mu_v).
\end{aligned}$$

The characteristic curve Γ_ν is therefore a quartic curve with a double point at 0, one nodal tangent always lying in the given direction μ, the other depending on ν. This second tangent lies in the direction μ if, and only if, $\nu = \mu$, and lies in the direction ν only if $\nu = \mu$ or $\nu = -1/\mu^2$. The osculating plane at 0 of that branch of Γ_ν which lies in the direction μ has the equation

$$24\mu^4(\mu^3 + 1)(\mu - \nu)(\mu X_2 - X_3) - (D + E\mu)X_4 = 0,$$

which reduces, when $\nu \neq \mu$, to (8). Hence, if $\nu \neq \mu$, the osculating plane at 0 of that nodal branch of the characteristic curve Γ_ν which lies in the direction μ is independent of ν and is the osculating plane of the pangeodesic in the direction μ.

If $\nu \neq \mu$, the osculating plane of the second nodal branch of Γ_ν at 0 is expressible in the form

$$\begin{aligned}
24\mu^4(\mu^3 + 1)(\mu - \nu)[(2\nu - 3\mu - \mu^3\nu)X_2 - (2\mu^3 - 3\mu^2\nu - 1)X_3] \\
+ [D(2\mu^3 - 3\mu^2\nu - 1) + E(2\nu - 3\mu - \mu^3\nu)]X_4 = 0.
\end{aligned}$$

As ν varies this plane generates an axial pencil, its limiting position as ν approaches μ being given by

$$\begin{aligned}
(41) \quad & 6\mu(\mu^3 + 1)(\mu X_2 - X_3) \\
& + [A\mu^2(11 - 4\mu^3) + B\mu(4 - 11\mu^3) + 9\mu^6 - 9 + 36(\mu^3 + 1)\mu']X_4 = 0,
\end{aligned}$$

where μ' is defined in (7). The plane (41) depends on the curve C_μ alone, being independent of the other curves of the family $C^{(1)}$, and together with the osculating plane of C_μ at 0, the osculating plane of the pangeodesic in the direction μ at 0, and the tangent plane $X_4 = 0$, determines an invariant cross-ratio of -4. Hence, the plane (41) coincides at every point of the surface S with the osculating plane of the curve C_μ if, and only if, the curves $C^{(1)}$ are pangeodesics.

If $\nu = \mu$, (40) reduces to

$$[X_4 - n_1(X_3 - \mu X_2)][X_4 - n_2(X_3 - \mu X_2)] = 0, \tag{42}$$

where n_1, n_2 are the roots of the equation

$$Fn^2 + En - 12\mu^4(\mu^3 + 1) = 0 \qquad (\nu = \mu). \tag{43}$$

The characteristic curve Γ_ν therefore consists of two non-degenerate conics determined by the intersections of the quadric Q_μ and the planes (42). The harmonic conjugate of the plane $X_4 = 0$ with respect to these two planes has the form

$$24\mu^4(\mu^3 + 1)(\mu X_2 - X_3) + EX_4 = 0 \qquad (\nu = \mu),$$

from which it follows that this plane, the osculating plane of C_μ at 0, the osculating plane of the pangeodesic in the direction μ at 0, and the plane $X_4 = 0$, determine a cross-ratio of -2.

The curves $C^{(1)}$ cannot be chosen so that at each point of S the two planes (42) coincide with the planes of sextactic section (28) which pass through the tangents to $C^{(1)}$. However, comparison of equations (43) and (23) shows that this situation does occur at a particular point 0 of S if, and only if, C_μ has second-order contact at 0 with any one of a set of twelve pangeodesics.

The quadric Q_μ touches the envelope E of the two-parameter family of Moutard quadrics $Q^{(2)}$ at the points determined by equations (9), (40) and (42), where the value of ν in (40) is arbitrary but different from μ. Since the two planes (42) meet the cone (40) in two lines l_1, l_2 which pass through 0 and are ordinarily distinct, and in the tangent $X_3/X_2 = \mu$, $X_4 = 0$, and since these three lines intersect Q_μ at 0 and at two other points P_1, P_2, it follows that *the envelope E consists of the surface S and of two other sheets; those generated by the points P_1, P_2.*

The lines l_1, l_2 coincide if, and only if,

$$E^2 + 48F\mu^4(\mu^3 + 1) = 0, \tag{44}$$

where $\nu = \mu$ in the expressions for E, F. Hence, at a particular point 0 of S there are two planes either of which may serve as the osculating plane of the curve C_μ in order that l_1 and l_2 coincide. It is impossible, however, to choose the curves $C^{(1)}$ so that the lines l_1, l_2 coincide for every point of S, since equation (35) which determines the curves $C^{(1)}$ cannot satisfy (44) which is now a differential equation of the second order and second degree in dv/du.

INDIANA UNIVERSITY,
BLOOMINGTON, INDIANA.

A GENERALIZATION OF LOCAL CLASS FIELD THEORY.*

By O. F. G. SCHILLING.[1]

In this paper we present a new generalization of local class field theory. The underlying ground fields will be perfect with respect to an archimedian valuation whose associated value group contains the additive group of all rational numbers and a finite number of infinite cycles which are generated by irrational real numbers. The resulting theory has many similarities with the theory of fields of formal power series of several variables.[2] We put stress on the investigation of the class group of algebras; it turns out that there exist in general division algebras which do not possess unramified splitting fields. Consequently, the theory has an entirely different aspect compared with the local class field theory as developed in earlier papers of M. Moriya and the author.[3] Only in two very special cases one can reestablish the classical results. We incorporated in the paper a result concerning the relationship between finite separable extensions and normal algebras over everywhere dense subfields.[4] Moreover, we prove the existence of perfect fields for which our general theory was developed. These and other examples seem to indicate that an axiomatic treatment of the algebraic structure of general perfect fields along the same lines as in the case of discrete perfect fields[5] will be quite involved, in particular the structure of the fields considered as additive groups is very complicated.

1. Preparations. Let k be an abstract field on which an archimedian valuation $\mathfrak{p}$ is defined, i. e. k is a field whose elements $a \neq 0$ can be mapped upon elements $v(a) = \alpha$ of a group $\Gamma(k)$ of real numbers; moreover, the function $v(a)$ satisfies the following conditions

---

* Presented to the American Mathematical Society April 16, 1938. Received by the Editors March 28, 1938.

[1] Johnston Scholar at the Johns Hopkins University for 1937-1938.

[2] O. F. G. Schilling, "Arithmetic in fields of formal power series in several variables," *Annals of Mathematics*, vol. 38 (1937).

[3] M. Moriya and O. F. G. Schilling, (1) "Zur Klassenkörpertheorie über unendlichen perfekten Körpern," *Journal of the Faculty of Science Hokkaido Imperial University Sapporo* (Japan), ser. I, vol. 5 (1936); (2) "Divisionsalgebren über unendlichen perfekten Körpern," *loc. cit.*, ser. I, vol. 6 (1937).

[4] O. F. G. Schilling, "A note on infinite perfect fields," *Abstract in the Bulletin of the American Mathematical Society*, vol. 44 (1938), no. 1, part 1.

[5] O. F. G. Schilling, "The structure of local class field theory," *American Journal of Mathematics*, vol. 60 (1938).

$$v(ab) = v(a) + v(b)$$
$$v(a + b) \geqq \min(v(a), v(b)).$$

In order to include the element 0 of k we define $v(0) = \infty$ and assume the usual formal relations to hold between the elements of $\Gamma(k)$ and the symbol ∞.

All elements a in k whose values are non-negative form the *maximal order* $\mathfrak{o}$ of all $\mathfrak{p}$-adic integers in k. The maximal order $\mathfrak{o}$ contains a *prime ideal* $\mathfrak{p}$ which is given by all elements of k whose values are greater than the zero-element of $\Gamma(k)$. The ring of residual classes $\mathfrak{o}/\mathfrak{p}$ is a field $\boldsymbol{k}$, the so-called *residue class field* of k with respect to the valuation $\mathfrak{p}$.[6]

We shall assume throughout the paper that the fields k and $\boldsymbol{k}$ have the same characteristic χ. However, most of the theorems we shall prove will also hold if we drop this assumption.

All elements of k form the points of a metric space if we define the distance $\delta(a, b)$ between two elements a and b of k to be equal to $C^{-\delta(a-b)}$ where C denotes any positive number greater than 1. The space k need not be complete with respect to the metric, but in any case we can complete it to the so-called *perfect $\mathfrak{p}$-adic closure* by adjoining all fundamental sequences which are defined by the metric.

It turns out that by completing the field k we enlarge neither the value group $\Gamma(k)$ nor the field of residual classes $\boldsymbol{k}$.

The essential feature of perfect $\mathfrak{p}$-adic fields k is expressed by Hensel's criterion of reducibility:[7]

If the polynomial $f(x) = x^n + a_1x^{n-1} + \cdots + a_n$ with integral coefficients in the perfect field k decomposes mod $\mathfrak{p}$ into the product $\prod_{i=1}^{m} \phi_i(x)$ of relatively prime polynomials $\phi_i(x)$ with coefficients in $\boldsymbol{k}$ then $f(x)$ already decomposes in k and $f(x) = \prod_{i=1}^{m} \Phi_i(x)$ where $\Phi_i(x)$ mod $\mathfrak{p} = \phi_i(x)$, $\phi_i(x)$ and $\Phi_i(x)$ have the same degrees.

It can be shown that there do exist fields k on which an archimedian valuation $\mathfrak{p}$ is defined such that Hensel's criterion holds for the polynomials with coefficients in the maximal order of k although the fields under consideration are not perfect. According to A. Ostrowski we shall term such fields *relatively perfect.* The validity of Hensel's criterion implies that the valuation

---

[6] For general statements about p-adic number fields see e. g. M. Deuring, *Algebren* (Berlin, 1935), Chap. VI, § 10. This tract will be quoted by D.

[7] A. Ostrowski, "Untersuchungen zur arithmetischen Theorie der Körper," *Mathematische Zeitschrift,* vol. 39 (1934).

$\mathfrak{p}$ of any relatively perfect field k possesses a unique extension $\mathfrak{P}(K)$ corresponding to each algebraic extension K of k. The value of an element $A \varepsilon K$ is given by $[K:k]^{-1}v(\mathfrak{N}A)$ where $\mathfrak{N}A$ denotes the norm of A taken from K to k. The integers of K are uniquely characterized by the fact that their norms are $\mathfrak{p}$-adic integers.

DEFINITION. *A subfield k' of a field k on which a valuation $\mathfrak{p}$ is defined is said to be everywhere dense in k if*

(I) *the valuation $\mathfrak{p}$ of k induces in k' a non-trivial valuation such that*

$$\Gamma(k) = \Gamma(k'),$$

(II) *the residue class fields* $\mathbf{k}$ *and* $\mathbf{k}'$ *coincide,*

(III) *k and k' are relatively perfect.*

THEOREM 1. *If k' is an everywhere dense subfield of k then every finite separable algebraic extension K of k is the join of a suitable extension K' of k' with k; moreover $[K:k] = [K':k']$.*

Proof. The fields k and k' obviously have the same $\mathfrak{p}$-adic closure $\bar{k}$. Let us assume that we proved the theorem for the pairs of fields $k, \bar{k}$ and $k', \bar{k}$. Then any field $\bar{K}$ over $\bar{k}$ has the form

$$\bar{K} = \bar{k} \cdot k(A) \text{ and}$$
$$\bar{K} = \bar{k} \cdot k'(A')$$

where $[\bar{K}:\bar{k}] = [A:k] = [A':k']$. Since also $[A':\bar{k}] = [A':k']$ we find

$$[A':k] = [A':k'] = [A:k].$$

Consequently

$$k(A) = k(A') = kk'(A').$$

Thus it suffices to prove the theorem under the assumption that k is a perfect $\mathfrak{p}$-adic field.

We consider the polynomials $f(x) = \sum_{i=0}^{n} a_i x^i$ of fixed degree n with coefficients as elements of a topological space $S_n(k)$ where we define

$$\delta(f(x), g(x)) = C^{-v(f(x)-g(x))}, \qquad C > 1$$

and

$$v(f(x)) = \min(v(a_0), \cdots, v(a_i), \cdots, v(a_n)).^{8}$$

The function $v(f(x))$ has the following properties

---

[8] F. K. Schmidt, "Mehrfach perfekte Körper," *Mathematische Annalen*, vol. 108 (1933).

$$v(f(x)+g(x)) \geqq \min(v(f(x)), v(g(x)))$$
$$v(f(x)g(x)) \geqq v(f(x)) + v(g(x)).$$

It can be proved that two polynomials $f(x)$, $g(x)$ in $S_n(k)$ have the same type of decomposition relative to k if $v(f(x) - g(x))$ is sufficiently large. We shall make use of this statement later on.

Let $g(x)$ be an arbitrary polynomial of $S_n(k)$ then

$$g(x+t) = g(x) + tD^{(1)}g(x) + t^2D^{(2)}g(x) + \cdots + t^iD^{(i)}g(x) + \cdots + t^nD^{(n)}g(x)$$

where the differential operator $D^{(i)}$ is defined by

$$D^{(i)}g(x) = \frac{d^ig(x)}{dx^i} \cdot \frac{1}{i!}.$$

It can be shown that such formal derivatives are also defined if the underlying field of coefficients k has a characteristic $\chi \neq 0$.[9]

Now let K be a finite separable algebraic extension of k. We always can assume that K is generated by a primitive element A such that its irreducible equation $f(x)$ has the form

$$a_0 + a_1x + \cdots + a_ix^i + \cdots + a_{n-1}x^{n-1} + x^n$$

where $a_0, \cdots, a_{n-1}$ are $\mathfrak{p}$-adic integers. Since $f(x)$ is a separable polynomial we have $D^{(1)}f(A) \neq 0$, moreover

$$v(D^{(1)}f(A)) = \mathrm{A} \geqq 0$$

for A is an integer of K.[10]

Now let a_i be an arbitrary coefficient of $f(x)$ then there exist elements a'_i in the everywhere dense subfield k' such that

$$v(a_i - a'_i) > \Delta > 2\mathrm{A} \geqq 0$$

where Δ is arbitrarily large but fixed. Let us then choose such a set of numbers a'_i and let us consider the polynomial

$$g(x) = \sum_{i=0}^{n} a_i'x^i$$

of degree n in k'. Then

$$v(f(x) - g(x)) > \Delta.$$

---

[9] H. Hasse, "Noch eine Begründung der Theorie der höheren Differentialquotienten in einem algebraischen Funktionenkörper einer Unbestimmten," *Crelle*, vol. 177 (1937).

[10] Cf. M. Moriya, "Klassenkörpertheorie im Kleinen für die unendlichen algebraischen Zahlkörper," *Journal of the Faculty of Science Hokkaido Imperial University Sapporo* (Japan), ser. I, vol. 5 (1936).

Since the coefficients a_i of $f(x)$ are integers the relation $v(f(x) - g(x)) > \Delta$ implies that the a_i' are integers of k'; $g(x)$ obviously is also a separable polynomial.

We have

$$v(g(A) - f(A)) = v(g(A)) > \Delta$$

when we substitute for the undeterminate x the root A.

Next we show that $D^{(1)}g(A) \neq 0$. Namely, $v(g(x) - f(x)) > \Delta$ implies that also $v(D^{(1)}g(x) - D^{(1)}f(x)) > \Delta$. Since A is an integer we can infer

$$v(D^{(1)}g(A) - D^{(1)}f(A)) > \Delta.$$

If we assume that $v(D^{(1)}g(A)) \neq v(D^{(1)}f(A))$ then

$$v(D^{(1)}g(A) - D^{(1)}f(A)) = \min(D^{(1)}g(A), D^{(1)}f(A)) > \Delta > \mathrm{A}$$

in contradiction to $v(D^{(1)}f(A)) = \mathrm{A} < \Delta$.

Hence

$$\infty > v(D^{(1)}f(A)) = v(D^{(1)}g(A)) = \mathrm{A}, \text{ or } D^{(1)}g(A) \neq 0.$$

Consequently we can solve the equation

$$g(A) + t_1 D^{(1)}g(A) = 0$$

with respect to t_1.

Since $v(g(A)) > \Delta > 2\mathrm{A}$ we can put

$$v(g(A)) = 2\mathrm{A} + \mathrm{E}$$

where $\mathrm{E} > 0$.

Then

$$v(t_1) = v(g(A)) - v(D^{(1)}g(A)) = \mathrm{A} + \mathrm{E} > \mathrm{A}$$

for

$$v(D^{(1)}g(A)) = \mathrm{A} < \Delta.$$

Consider next the integral element

$$A_1 = A + t_1$$

and form

$$g(A_1) = g(A + t_1) = g(A) + t_1 D^{(1)}g(A) + \cdots + t_1^n D^{(n)}g(A).$$

The coefficients of the derivatives $D^{(i)}g(x)$ are easily seen to be integral. Consequently $D^{(i)}g(A)$ are integers for A was assumed to be an integer over k. Therefore we get

$$v(t_1^i D^{(i)}g(A)) \geqq i v(t_1) \geqq i(\mathrm{A} + \mathrm{E}) \geqq 2(\mathrm{A} + \mathrm{E})$$

provided that $n \geqq i \geqq 2$. Hence

$$v(g(A_1)) = 2(\mathrm{A} - \mathrm{E}).$$

Putting $v(g(A_1)) = 2\mathrm{A} + \mathrm{E}_1$ we have

$$\mathrm{E}_1 \geqq 2\mathrm{E} \quad \text{and} \quad v(g(A_1)) = 2\mathrm{A} + \mathrm{E}_1 > 2\mathrm{A} + \mathrm{E} = v(g(A)).$$

We find

$$v(D^{(1)}f(A_1)) = v(D^{(1)}f(A + t_1)) = \mathrm{A}$$

for $v(D^{(1)}f(A)) = \mathrm{A}$ and $v(t_1) > \mathrm{A}$.

As before we can prove that

$$v(D^{(1)}g(A_1)) = \mathrm{A},$$

hence

$$D^{(1)}g(A_1) \neq 0,$$

and we can determine t_2 from

$$g(A_1) + t_2 D^{(1)}g(A_1) = 0.$$

Moreover,

$$\begin{aligned} v(t_2) &= v(g(A_1)) - v(D^{(1)}g(A_1)) \\ &= 2\mathrm{A} + \mathrm{E}_1 - \mathrm{A} = \mathrm{A} - \mathrm{E}_1 \geqq \mathrm{A} + 2\mathrm{E}. \end{aligned}$$

Introducing the integer $A_2 = A_1 + t_2$ we find

$$v(g(A_2)) \geqq 2(\mathrm{A} - \mathrm{E}_1);$$

if we put

$$v(g(A_2)) = 2\mathrm{A} + \mathrm{E}_2,$$

then

$$\mathrm{E}_2 \geqq 2\mathrm{E}_1 \geqq 2^2\mathrm{E}.$$

Again

$$v(D^{(1)}g(A_2)) = \mathrm{A}.$$

Repeating these arguments we can construct an element t_i such that

$$v(t_i) \geqq \mathrm{A} + 2^{i-1}\mathrm{E} \quad \text{and} \quad v(g(A_i)) = 2\mathrm{A} + \mathrm{E}_i \geqq 2\mathrm{A} + 2^i\mathrm{E}$$

where

$$A_i = A_{i-1} + t_i.$$

Let us consider the sequence of elements

$$A, A_1 = A + t_1, \cdots, A_i = A + t_1 + \cdots + t_i, \cdots.$$

These elements form a fundamental sequence of algebraic quantities over k for

$$v(A_i - A_{i+j}) \geqq \mathrm{A} + 2^i\mathrm{E}$$

if $j \geqq 1$.

Since K is a perfect field the sequence $\{A_i\}$ has a limit $\bar{A}$ in K.

According to our construction we have

$$\lim_{i\to\infty} \nu(g(A_i)) = \infty,$$

hence the sequence $\{g(A_i)\}$ converges to 0 in K, or ultimately

$$g(\bar{A}) = 0 \text{ in } K.$$

Since the polynomial $g(x)$ is chosen in k' the root $\bar{A}$ of $g(x) = 0$ is algebraic over k'. Moreover, $g(x)$ is irreducible in k according to the remark at the beginning of the proof; a fortiori $g(x)$ is irreducible in k'. Consequently $[\bar{A} : k'] = [\bar{A} : k] = [K : k]$ or

$$K = k(\bar{A}) = kk'(\bar{A}).$$

In particular, $k'(\bar{A})$ is a normal extension of k' if K is normal over k'.

The theorem we just proved has an analogue for the normal simple algebras over k and k'.

THEOREM 2. *If k' is an everywhere dense subfield of k then each normal algebra $\mathfrak{A}$ over k is similar to the direct product of a normal simple algebra $\mathfrak{A}'$ over k' with the field k.*

Proof. Let $\mathfrak{D}$ be an arbitrary normal division algebra over k. Then $\mathfrak{D}$ can be represented uniquely as the direct product $\mathfrak{D}' \times \mathfrak{D}^{(\chi)}$ of a division algebra $\mathfrak{D}'$ whose degree is relatively prime to the characteristic χ and another division algebra $\mathfrak{D}^{(\chi)}$ whose degree is a power of χ. Consequently, it suffices to prove the theorem for the components separately. As in Theorem 1 it suffices to prove the theorem in the case that k is perfect.

Case i. $\mathfrak{D} = \mathfrak{D}'$.

The division algebra $\mathfrak{D}$ can be represented by a crossed product $(a_{\sigma,\tau}, K/k)$ where K denotes a suitable normal extension of k whose Galois group consists of the elements $\{\sigma, \tau, \cdots\}$. Since the degree of $\mathfrak{D}$ is relatively prime to χ we have

$$a^e_{\sigma,\tau} = c_\sigma c_\tau{}^\sigma c_{\sigma\tau}^{-1}$$

where $(e, \chi) = 1$. Consequently there exists the least normal extension N over k which contains K and the quantities $c_\sigma{}^{1/e}, c_\tau{}^{1/e}, \cdots$. Let $\{\Sigma, \mathrm{T}, \cdots\}$ be the Galois group of N/k then

$$\mathfrak{D} \sim (a_{\sigma,\tau}, K/k) \sim (\zeta^{(e)}_{\Sigma,\mathrm{T}}, N/k)$$

where $\zeta^{(e)}_{\Sigma,\mathrm{T}}$ are suitable e-th roots of unity.[11] Now we can apply Theorem 1, the field N has the form $N = N' \cdot k$ where $[N' : k'] = [N : k]$ and where the respective Galois groups are isomorphic. Observing that the quantities $\zeta^{(e)}_{\Sigma,\mathrm{T}}$ are absolutely algebraic, we can construct the algebra

---

[11] D. Chap. V, § 7.

$$(\zeta^{(e)}_{\Sigma,\mathrm{T}}, N'/k') \text{ over } k',$$

obviously

$$\mathfrak{D} \sim (\zeta^{(e)}_{\Sigma,\mathrm{T}}, N'/k') \times k.$$

Case ii. $\mathfrak{D} = \mathfrak{D}^{(\chi)}$.

According to a theorem of A. A. Albert each division algebra of the type we have to consider, is similar to the direct product of cyclic crossed products. Consequently, it suffices to prove the theorem for cyclic algebras $(a, Z/k)$. Let Z' be the separable extension of k' such that $Z = Z'k$. We have to construct a factor set a' in k' such that

$$(a', Z'/k') \times k \sim (a, Z/k).$$

Since k was assumed to be perfect, the element a of k is the limit of a fundamental sequence $\{a'_i\}$ of elements a'_i in k'. Using the fundamental properties on the convergence of $\mathfrak{p}$-adic sequences and series, we find that

$$a = \sum_{i=i_0}^{\infty} (a'_i - a'_{i-1})$$

where i_0 denotes the least subscript such that

$$v(a'_{i_0}) = v(a).$$

The section $a' = \sum_{i=i_0}^{m} (a'_i - a'_{i-1})$ denotes an element of k' whose value $v(a')$ coincides with $v(a)$. The element $aa'^{-1} = b$ is a unit of k and

$$b = 1 + c$$

where $v(c) > 0$. The usual arguments concerning $\mathfrak{p}$-adic valuations yield that for any prescribed $N > 0$ in the value group $\Gamma(k)$ there exists a number $m = m(N)$ *such that*

$$v(c) \geqq N.$$

Now let us consider the expression

$$1 + y = \mathfrak{N}(1 + xA) = 1 + t(A)x + \cdots$$

where A denotes a suitable primitive element of Z/k whose trace $t(A)$ does not vanish. The element x can be expressed in terms of a convergent power series in y provided that $v(y)$ is sufficiently large, say $v(y) \geqq N$. Consequently the element $1 + y$ is a norm. We apply this result to the unit $b = 1 + c$

where we assume that $v(c) \geqq N$, i.e. that m is sufficiently large. Then $b = \mathfrak{N}B$ with $B \varepsilon Z$, hence

$$(a, Z/k) = (a'b, Z/k) \sim (a', Z/k) \times (b, Z/k)$$
$$\sim (a', Z/k).$$

Again the field Z has the form $Z'k$ where Z' is cyclic over k', hence obviously

$$(a', Z'/k') \times k \sim (a, Z/k).$$

Remark. In Theorems 1 and 2 we can omit the assumption that the subfield k' be relatively perfect.

2. Specialization of the value group $\Gamma(k)$. Let K be an arbitrary normal extension of the relatively perfect field k and let $G = \{\sigma, \tau, \cdots\}$ be its Galois group. Since k is assumed to be relatively perfect the uniquely determined prime ideal $\mathfrak{P}$ of K is left invariant by all substitutions of the Galois group. G contains a set of subgroups which are determined by the arithmetical properties of the extension K.[12]

I) all substitutions $\sigma \varepsilon G$ such that

$$\epsilon^\sigma = \epsilon\eta \text{ with } \eta \equiv 1 \pmod{\mathfrak{P}}$$

for all units ϵ of K, form an invariant subgroup I of G, the so-called *inertial group.*

II) All substitutions $\tau \varepsilon I$ such that

$$a^\tau = a\eta \text{ with } \eta \equiv 1 \pmod{\mathfrak{P}}$$

holds for all elements a of K form an invariant subgroup R of G the so-called *ramification group.*

The fields $K(I)$ and $K(R)$, which belong to the groups I and R respectively, are called the inertial and ramification fields respectively. If we assume that the degree $[K:k]$ is relatively prime to the characteristic χ then it can be proved that $K(I)$ is the maximal subfield U of K such that

$$\Gamma(U) = \Gamma(k) \text{ and } \boldsymbol{U} = \boldsymbol{K}, \text{ i.e.}$$

$K(I)$ is the maximal unramified extension lying in K, when an unramified extension U of k is characterized by the relations

$$\Gamma(U) = \Gamma(k) \text{ and } [U:k] = [\boldsymbol{U}:\boldsymbol{k}].$$

If U is in particular a normal extension of k then its Galois group induces the Galois group of $\boldsymbol{U}$ over $\boldsymbol{k}$.

---

[12] W. Krull, "Allgemeine Bewertungstheorie," *Crelle*, vol. 167 (1931).

The inertial field $K(I)$ is in any case uniquely determined as the maximal unramified subfield if the field of residue classes $\boldsymbol{k}$ is algebraically perfect.

The ramification field $K(R)$ is always an *abelian* extension of K whose Galois group I/R is isomorphic with the factor group $\Gamma(K(R))/\Gamma(K(I))$. If K is a proper algebraic extension of $K(R)$ then $[K:K(R)]$ is a power of the characteristic. Consequently $K(R)=K$ if $([K:k],\chi)=1$, and hence K is an abelian extension of $K(I)$ where

$$I\cong\Gamma(K)/\Gamma(K(I))=\Gamma(K)/\Gamma(k).$$

LEMMA 1. *If the residue class field $\boldsymbol{k}$ of the relatively perfect field k is algebraically closed then each extension K of k whose degree is relatively prime to the characteristic χ is abelian.*

Proof. Let K be an arbitrary algebraic extension of k whose degree is relatively prime to χ. The field K is contained in a least normal extension N of k. Since $\boldsymbol{k}$ is assumed to be algebraically closed we have $N(I)=k$. The results concerning the ramification field $N(R)$ imply that

$$K\subseteq N(R)$$

for $([K:k],\chi)=1$ and $[N:N(R)]$ is at most a power of χ. Consequently, K is an abelian extension of k for the Galois group of $N(R)$ over k is isomorphic with $\Gamma(N(R))/\Gamma(k)\cong\Gamma(K)/\Gamma(k)$.

We now wish to specialize the value group $\Gamma(k)$ in order to be able to obtain more precise results concerning the theory of algebras which we have to develop later on. We assume that $\Gamma(k)$ contains the additive group of all rational numbers P and a finite number of infinite cyclic groups $\theta_i\Gamma^{(0)}$ where θ_i $(i=1,\cdots,\lambda)$ denote a set of rationally independent real numbers and where $\Gamma^{(0)}$ denotes the additive group of all integers:

$$\Gamma(k)=\Gamma=\{\mathrm{P},\theta_1\Gamma^{(0)},\cdots,\theta_\lambda\Gamma^{(0)}\}.$$

Each element $\alpha\in\Gamma$ can then uniquely be represented in the form

$$\alpha=\beta+\sum_{i=1}^{\lambda}\gamma_i\theta_i$$

where $\beta\in\mathrm{P}$ and $\gamma_i\in\Gamma^{(0)}$.

LEMMA 2. *Let k^* and $\boldsymbol{k}^*$ denote the multiplicative groups of k and $\boldsymbol{k}$ respectively. If the integer n is relatively prime to the characteristic χ of k then*

$$[k^*:k^{*n}]=[\Gamma/n\Gamma:1][\boldsymbol{k}^*:\boldsymbol{k}^{*n}]$$

if $[\boldsymbol{k}^*:\boldsymbol{k}^{*n}]$ is finite.

Proof. Let us fix λ elements $t_1, \cdots, t_\lambda$ in k such that

$$v(t_i) = \theta_i \qquad (i = 1, \cdots, \lambda).$$

Then each element a of k^* can be written in the form

$$a = \prod_{i=1}^{\lambda} t_i^{\gamma_i} b$$

where

$$v(a) = \beta + \sum_{i=1}^{\lambda} \gamma_i \theta_i \quad \text{and} \quad v(b) = v(a \prod_{i=1}^{\lambda} t_i^{-\gamma_i}) = \beta \,\varepsilon\, \mathrm{P}.$$

All elements $b \,\varepsilon\, k^*$ whose values $v(b)$ lie in the rational component P form a multiplicative subgroup B of k^*, and obviously $k^* = \langle \{t_1\}, \cdots, \{t_\lambda\}, B \rangle$ where $\{t_i\}$ denote the infinite cycles generated by the elements t_i; $\langle \ldots \rangle$ denotes the join of $\{t_1\}, \cdots, \{t_\lambda\}$ and B. Since 1 is the only element of k^* contained in both $\langle \{t_1\}, \cdots, \{t_\lambda\} \rangle$ and B we see that

$$k^* = \langle \{t_1\}, \cdots, \{t_\lambda\} \rangle \times B.$$

We find $k^{*n} = \langle \{t_1\}^n, \cdots, \{t_\lambda\}^n \rangle \times B^n$ as a consequence of

$$a^n = \prod_{i=1}^{\lambda} t_i^{\gamma_i n} b^n.$$

Consequently,

$$\begin{aligned} k^*/\langle \{t_1\}, \cdots, \{t_\lambda\} \rangle^n \times B &\cong \{t_1\}/\{t_1\}^n \times \cdots \times \{t_\lambda\}/\{t_\lambda\}^n \times B/B \\ &\cong \{t_1\}/\{t_1\}^n \times \cdots \times \{t_\lambda\}/\{t_\lambda\}^n. \end{aligned}$$

We have

$$\begin{aligned} [k^* : k^{*n}] &= [k^*/\langle \{t_1\}, \cdots, \{t_\lambda\} \rangle^n \times B : 1][\langle \{t_1\}, \cdots, \{t_\lambda\} \rangle^n \times B/k^{*n} : 1] \\ &= n^\lambda [\langle \{t_1\}, \cdots, \{t_\lambda\} \rangle^n \times B/k^{*n} : 1] \\ &= n^\lambda [\langle \{t_1\}, \cdots, \{t_\lambda\} \rangle^n \times B/\langle \{t_1\}, \cdots, \{t_\lambda\} \rangle^n \times B^n : 1] \\ &= n^\lambda [\langle \{t_1\}, \cdots, \{t_\lambda\} \rangle^n \times B/\langle \{t_1\}, \cdots, \{t_\lambda\} \rangle^n/\langle \{t_1\}, \cdots, \{t_\lambda\} \rangle^n \\ &\quad \times B^n/\langle \{t_1\}, \cdots, \{t_\lambda\} \rangle^n : 1] \\ &= n^\lambda [B/B^n : 1]. \end{aligned}$$

We next have to reduce the index $[B/B^n : 1]$ under the assumption that $(n, \chi) = 1$. Let b be an arbitrary element of B, then $v(b) = \beta \epsilon \mathrm{P}$. Consequently, there exists at least one element c in k such that $v(c) = \beta/n$. Therefore $b = c^n \epsilon$ where ϵ is a suitable unit. Hence $B = \langle B^n, \epsilon \rangle$ where ϵ stands for the group of units in k. Therefore

$$\begin{aligned} B/B^n = \langle B^n, \epsilon \rangle / \langle B^n, \epsilon^n \rangle &= \langle B^n, \epsilon \rangle / B^n \\ &\cong \epsilon / B^n \cap \epsilon = \epsilon/\epsilon^n \end{aligned}$$

since $v(\epsilon) = 0$.

Now let η be a unit of k and let $\boldsymbol{a}$ be its uniquely determined residue class mod $\mathfrak{p}$. If we assume that the equation $x^n - \boldsymbol{a} = 0$ is divisible in $\boldsymbol{k}$ by a linear factor $x - \boldsymbol{b}$, i. e. $x^n - \boldsymbol{a} = (x - \boldsymbol{b})\boldsymbol{g}(x)$, such that $((x - \boldsymbol{b}), \boldsymbol{g}(x)) = 1$ then the polynomial $x^n - \eta = 0$ is divisible by a linear factor $x - \eta'$ where $\eta' \bmod \mathfrak{p} = \boldsymbol{b}$ for Hensel's criterium of reducibility holds. Consequently, each unit η which is $\equiv 1 \pmod{\mathfrak{p}}$ can be written as a n-th power.

Now let us consider $\mathfrak{p}$ as a multiplicative mapping of the group of all units ϵ upon the multiplicative group $\boldsymbol{k}^*$. Then

$$[\epsilon/\epsilon^n : 1] = [\boldsymbol{k}^* : \boldsymbol{k}^{*n}].$$

as the application of Herbrand's Lemma yields. Hence finally

$$[k^* : k^{*n}] = n^\lambda[\boldsymbol{k}^* : \boldsymbol{k}^{*n}].$$

Remark. If we assume that the residue class field $\boldsymbol{k}$ possesses for each integer n which is relatively prime to χ, exactly one cyclic extension $\boldsymbol{U}_n$ of degree n then obviously

$$[\boldsymbol{k}^* : \boldsymbol{k}^{*n}] \leqq n.$$

Namely, if $\boldsymbol{k}$ contains the n-th roots of unity then $\boldsymbol{U}_n = \boldsymbol{k}(\boldsymbol{a}^{1/n})$ where $\boldsymbol{k}^* = \boldsymbol{k}^{*n} + \boldsymbol{a}\boldsymbol{k}^{*n} + \cdots + \boldsymbol{a}^{n-1}\boldsymbol{k}^{*n}$, and if $\boldsymbol{k}$ does not contain the n-th roots of unity then surely $[\boldsymbol{k}^* : \boldsymbol{k}^{*n}] \leqq n$ for otherwise there would exist more than one cyclic extension $\boldsymbol{U}_n$ as a simple application of the Galois theory shows. Therefore we have

$$[k : k^{*n}] \leqq n^{\lambda+1}.$$

The structure of k^*/k^{*n} is given by

$$\{t_1\} \times \cdots \times \{t_\lambda\}/\{t_1\}^n \times \cdots \times \{t_\lambda\}^n \times \boldsymbol{k}^*/\boldsymbol{k}^{*n} \cong \Gamma(k)/n\Gamma(k) \times \boldsymbol{k}^*/\boldsymbol{k}^{*n}.$$

COROLLARY. *Let K be an arbitrary algebraic extension of k whose degree is relatively prime to χ then obviously $k^{*n} \subseteq \mathfrak{N}K^*$ where $\mathfrak{N}K^*$ denotes the multiplicative group of the norms of all elements in K^*. Then obviously*

$$[k^*/\mathfrak{N}K^* : 1] \leqq n^{\lambda+1}$$

as the application of the preceding remark shows.

LEMMA 3. *If the field of residue classes $\boldsymbol{k}$ is algebraically closed then the largest abelian extension K of k which has exponent n — $(n, \chi) = 1$—is given by $k(t_1^{1/n}, \cdots, t_\lambda^{1/n})$.*

Proof. According to the theory of radical fields we find that the largest abelian extension K of exponent n is given by adjoining the n-th roots of a set of representatives of k^*/k^{*n}. Hence $[K:k]=n^{\lambda}$ according to Lemma 2. Now $t_1,\cdots,t_{\lambda}$ form a complete set of representatives of k^*/k^{*n}, consequently $K=k(t_1^{1/n},\cdots,t_{\lambda}^{1/n})$. Incidentally, we find that each element b of k^* whose value is rational is a n-th power.

We next wish to investigate the structural properties of the class group of all normal simple algebras $\mathfrak{G}(k)$ over k. Now each division algebra $\mathfrak{D}$ with k as center is the direct product of a division algebra $\mathfrak{D}'$ whose degree is relatively prime to χ and of a division algebra $\mathfrak{D}^{(\chi)}$ whose degree is a power of χ. Consequently, the group $\mathfrak{G}(k)$ is the direct product of two subgroups $\mathfrak{G}'(k)$ and $\mathfrak{G}^{(\chi)}(k)$ where all classes in $\mathfrak{G}'(k)$ can be represented by division algebras of the type $\mathfrak{D}'$ and all algebras in $\mathfrak{G}^{(\chi)}(k)$ are representable by division algebras of type $\mathfrak{D}^{(\chi)}$. Therefore it suffices to investigate the structural properties of the two components.

3. The structure of $\mathfrak{G}^{(\chi)}(k)$. In order to describe the structure of $\mathfrak{G}^{(\chi)}(k)$ we can apply the results of A. A. Albert and other authors concerning χ-algebras.[13] With the help of this additive theory one proves

THEOREM 3. *Each division algebra in $\mathfrak{G}^{(\chi)}(k)$ is similar to the direct product of cyclic algebras.*

In general one cannot say more about the structure of $\mathfrak{G}^{(\chi)}(k)$ since it depends on the structure of k as an additive group, i. e. it is closely related to the invariantive characterization of general perfect fields. However, if we assume that the field of residue classes $\boldsymbol{k}$ admits for each degree χ^{ν} $(\nu=1,2,\cdots)$ exactly one cyclic extension $\boldsymbol{U}_{\nu}$ we can say more about $\mathfrak{G}^{(\chi)}(k)$.

THEOREM 4. *$\mathfrak{G}^{(\chi)}(k)$ contains a subgroup $\mathfrak{G}_u{}^{(\chi)}(k)$ which consists of the totality of all classes which are split by unramified fields U_{ν} $(\nu=1,2,\cdots)$. $\mathfrak{G}_u{}^{(\chi)}(k)$ is isomorphic with the direct sum of λ additive groups of type $\Gamma^{(0)}\chi^{-\infty}$ mod 1; where $\Gamma^{(0)}\chi^{-\infty}$ mod 1 denotes the additive group of all fractions* mod 1 *whose denominators are powers of χ.*

Proof. Let $U^{(\chi)}$ be the infinite unramified extension of k containing all the unramified cyclic fields U_{ν} which are uniquely determined by the extensions $\boldsymbol{U}_{\nu}$ over $\boldsymbol{k}$. If $\{\mathfrak{D}_1\}$ and $\{\mathfrak{D}_2\}$ are any two classes of $\mathfrak{G}^{(\chi)}(k)$ which are split by an unramified field then obviously

---

[13] A. A. Albert, "Normal division algebras of degree p^e over fields of characteristic p," *Transactions of the American Mathematical Society*, vol. 39 (1936); T. Nakayama, "Über Algebren über einem Körper von der Primzahlcharakteristik, II," *Proceedings of the Imperial Academy of Tokyo*, vol. 12 (1937).

$$\{\mathfrak{D}_1\} \times U^{(\chi)} \sim \{\mathfrak{D}_2\} \times U^{(\chi)} \sim U^{(\chi)}$$

whence

$$\{\mathfrak{D}_1\} \times \{\mathfrak{D}_2\} \times U^{(\chi)} \sim U^{(\chi)}$$

and

$$\{\mathfrak{D}_i^{-1}\} \times U^{(\chi)} \sim U^{(\chi)}$$

i. e. all algebras which have unramified splitting fields form a group $\mathfrak{G}_u^{(\chi)}(k)$.

In order to determine the structure of $\mathfrak{G}_u^{(\chi)}(k)$ it suffices to describe the structure of the subgroup consisting of all algebras which are split by a fixed unramified field U_ν for the fields U_ν form a linearly ordered tower of successive algebraic extensions with the limit $U^{(\chi)}$. The structure of these groups $\mathfrak{G}_{U_\nu}^{(\chi)}(k)$ is determined as follows. Any algebra $\mathfrak{A}$ in $\mathfrak{G}_{U_\nu}^{(\chi)}(k)$ can be represented as a cyclic crossed product $(a, U_\nu/k)$. Let us fix elements t_i $(i = 1, \cdots, \lambda)$ in k such that $v(t_i) = \theta_i$, then

$$v(a) = \alpha = \beta + \sum_{i=1}^{\lambda} \gamma_i \theta_i \text{ with } \beta \in \mathrm{P}$$

implies

$$a = b^{\chi^\nu} \prod_{i=1}^{\lambda} t_i^{\gamma_i} \epsilon$$

where b denotes an arbitrary element in k whose value $v(b)$ is equal to $\beta\chi^{-\nu}$. Hence

$$\begin{aligned}(a, U_\nu/k) &\sim (b^{\chi^\nu}, U_\nu/k) \times (t_1, U_\nu/k)^{\gamma_1} \times \cdots \times (t_\lambda, U_\nu/k)^{\gamma_\lambda} \times (\epsilon, U_\nu/k) \\ &\sim (t_1, U_\nu/k)^{\gamma_1} \times \cdots \times (t_\lambda, U_\nu/k)^{\gamma_\lambda} \times (\epsilon, U_\nu/k).\end{aligned}$$

We next prove that $(\epsilon, U_\nu/k) \sim k$.

First we remark that there do not exist cyclic algebras $(\epsilon, U_\nu/k)$ whose exponent is equal to the degree. Namely, suppose that the exponent and degree of $(\epsilon, U_\nu/k)$ coincide. Then $(\epsilon, U_\nu/k) = \mathfrak{D}$ is a division algebra. Consequently, the valuation $\mathfrak{p}$ of the ground field k can be extended to the algebra $\mathfrak{D}$ as we shall prove later on page 691. Hence the algebra of residue classes belonging to $\mathfrak{D}$ is a division algebra $\boldsymbol{D}$ over $\boldsymbol{k}$. Since the maximal order of $\mathfrak{D}$ contains the order which is given by all sums $\Sigma u^i \omega_i$ where ω_i are integers in U_ν, it follows that $\boldsymbol{D}$ has the center $\boldsymbol{k}$ and is explicitly given as $(\epsilon \bmod \mathfrak{p}, \boldsymbol{U}_\nu/\boldsymbol{k})$. Such a division algebra $\boldsymbol{D}$ cannot exist since we suppose that the field of residue classes $\boldsymbol{k}$ is algebraically perfect. Consequently $\mathfrak{D} \sim k$. Assume now that $(\epsilon, U_\nu/k)$ is an arbitrary algebra whose factor set is a unit ϵ of k. Suppose that the exponent s_1 of $(\epsilon, U_\nu/k)$ is greater than 1. We already proved that s_1 has to be a proper divisor of $[U_\nu : k]$. We have

$$(\epsilon^{s_1}, U_\nu/k) \sim (\epsilon, U_\nu^{(1)}/k) \sim k$$

where $U_\nu^{(1)}$ denotes the uniquely determined cyclic subfield of U_ν such that $[U_\nu : U_\nu^{(1)}] = s_1$. Consequently, $\epsilon = \mathfrak{N}_1\epsilon_1$ where ϵ_1 is a suitable unit of $U_\nu^{(1)}$ and where $\mathfrak{N}_1$ denotes the norm taken from $U_\nu^{(1)}$ to k. Next consider the algebra $(\epsilon_1, U_\nu/U_\nu^{(1)})$. Its exponent s_2 has to be a proper divisor of the degree $[U_\nu : U_\nu^{(1)}]$ as a repetition of the argument at the beginning of the proof shows, $U_\nu^{(1)}$ is also algebraically perfect. If s_2 should be greater than 1, then

$$(\epsilon_1^{s_2}, U_\nu/U_\nu^{(1)}) \sim (\epsilon_1, U_\nu^{(2)}/U_\nu^{(1)}) \sim U_\nu^{(1)}$$

where $U_\nu^{(2)}$ is the uniquely determined subfield of U_ν such that $[U_\nu : U_\nu^{(2)}] = s_2$. Hence $\epsilon_1 = \mathfrak{N}_2\epsilon_2$ where ϵ_2 is a unit of $U_\nu^{(2)}$ and where $\mathfrak{N}_2$ denotes norm taken from $U_\nu^{(2)}$ to $U_\nu^{(1)}$. We can repeat this argument only a finite number of times since the degree $[U_\nu : k]$ is finite. Observing that the norm is a transitive multiplicative function we find that $\epsilon = \mathfrak{N}E$ with a suitable unit E of U_ν. This statement contradicts the assumption that $(\epsilon, U_\nu/k) \nsim k$, therefore we must have $(\epsilon, U_\nu/k) \sim k$ in general.

Consequently, each algebra in $\mathfrak{B}_{U_\nu}{}^{(\chi)}(k)$ is similar to the direct product of λ algebras $(t_i, U_\nu/k)^{\gamma_i}$. The elements t_i are not norms of elements in U_ν for U_ν is an unramified extension of k. Hence

$$k^*/\mathfrak{N}U^*_\nu \cong \Gamma/\chi^\nu\Gamma \cong Z_1(\chi^\nu) \times \cdots \times Z_\lambda(\chi^\nu),$$

where $Z_i(\chi^\nu)$ denote λ cyclic groups of order χ^ν. Since each of the groups $Z_i(\chi^\nu)$ is isomorphic with the additive group of all fractions mod 1 whose denominators are divisors of χ^ν, we find for $\nu \rightarrow \infty$ that $\mathfrak{B}_{U_\nu}{}^{(\chi)}(k)$ is isomorphic with the direct sum of λ additive groups $\Gamma^{(0)}\chi^{-\infty}$ mod 1 consisting of all rational numbers mod 1 whose denominators are powers of χ.

The group $\mathfrak{B}_u{}^{(\chi)}(k)$ is in general a proper subgroup of $\mathfrak{B}^{(\chi)}(k)$ and it is not a direct factor of $\mathfrak{B}^{(\chi)}(k)$ as can easily be seen by examples. We wish to show by two examples that $\mathfrak{B}_u{}^{(\chi)}(k) \subset \mathfrak{B}^{(\chi)}(k)$; these examples will also illustrate how in general the additive structure and algebraic structure of the given ground field k influence the structure of $\mathfrak{B}^{(\chi)}(k)$. It is obvious that $\mathfrak{B}_u{}^{(\chi)}(k) \subset \mathfrak{B}^{(\chi)}(k)$ if $\lambda \geqq 2$. Namely, the algebra (t_2, t_1) with the generating relations

$$u^\chi = t_2,\ y^\chi - y = t_1,\ u^{-1}yu = y + 1$$

is not split if we extend k to the maximal unramified field $U^{(\chi)}$. In order to prove $\mathfrak{B}^{(\chi)}(k) \supset \mathfrak{B}_u{}^{(\chi)}(k)$ for a field k it is necessary and sufficient to show that $\mathfrak{B}^{(\chi)}(U) \neq 1$ where $U^{(\chi)}$ denotes the maximal unramified field over k. Let us assume for the present that $k = U^{(\chi)}$. If one considers infinite algebraic extensions of fields of formal power series of one variable with coefficients in an algebraically closed field of characteristic $\chi \neq 0$ then one finds that $\mathfrak{G}^{(\chi)}(k) = 1$. Thus one could suspect that $\mathfrak{G}^{(\chi)}(k) = 1$ if $\Gamma(k) = P$, or $= \{P, \theta\Gamma^{(0)}\}$. We shall give two examples of fields k whose value groups $\Gamma(k)$ have either one of the types mentioned and whose fields of residue classes are algebraically closed but for which $\mathfrak{G}^{(\chi)}(k) \neq 1$.

Let Ω be an arbitrary algebraically closed field of characteristic $\chi \neq 0$. Consider then the field $\Omega\{t\}$ which consists of all formal power series in t with coefficients in Ω. According to well-known results the field $\Omega\{t\}$ is perfect with respect to a discrete archimedian valuation $\mathfrak{p}$ and Ω is the associated field of residue classes. We adjoin to $\Omega\{t\}$ all elements $t^{1/n}$ where $(n, \chi) = 1$, the resulting field $\Omega\{t\}(\langle t^{1/n}\rangle)$ is an infinite algebraic extension of $\Omega\{t\}$ whose value group $\Gamma(\Omega\{t\}(\langle t^{1/n}\rangle))$ consists of all rational numbers whose denominators are relatively prime to χ. Again Ω is the field of residue classes. Next we adjoin to $\Omega\{t\}$ all solutions of the cyclic equations of degree χ^ν which are given by the roots of

$$\dot{y}_\nu{}^{\chi} \dot{-} \dot{y}_\nu = \dot{a}_\nu$$

where $\dot{y}_\nu = (y_1, \cdots, y_\nu)$ and $\dot{a}_\nu = (t^{-1}, \cdots, t^{-1})$.[14] The resulting cyclic field Z_ν has degree χ^ν over $\Omega\{t\}$ and $\mathfrak{p}$ is completely ramified in Z_ν, i. e. $\mathfrak{p} = \mathfrak{p}_\nu{}^{\chi^\nu}$ where $\mathfrak{p}_\nu$ denotes the prime ideal of Z_ν. The fields Z_ν form an infinite tower of relatively cyclic fields

$$\Omega\{t\} < Z_1 < \cdots < Z_i < Z_{i+1} < \cdots,$$

the join of these fields is an infinite cyclic extension Z_∞ of $\Omega\{t\}$ whose degree is equal to χ^∞. The value group $\Gamma(Z_\infty)$ consists of all rational numbers whose denominators are powers of χ, again Ω is the field of residue classes. The join of $\Omega\{t\}(\langle t^{1/n}\rangle)$ and Z_∞ is an infinite separable abelian extension A_∞ of $\Omega\{t\}$ whose value group $\Gamma(A_\infty)$ consists of all rational numbers P and whose field of residue classes $\mathfrak{o}_\infty/\mathfrak{p}_\infty$ is equal to the algebraically closed field Ω.

The equation $f(x) \equiv x^\chi - x = t^{-(1+\chi)} = 0$ is not solvable in any one of the fields $\Omega\{t\}$, $\Omega\{t\}(\langle t^{1/n}\rangle)$ and Z_∞, consequently it is not solvable in A_∞. Hence a solution of $f(x) = 0$ generates a cyclic extension Z of degree χ over A_∞, Z is ramified over A_∞. We only have to prove that $f(x)$ is not solvable in Z_∞. Consider for this purpose the discriminant of the field $\Omega\{t\}(x)$, it is equal $\mathfrak{p}^{(\chi-1)(2+\chi)}$ whereas the field $\Omega\{t\}(y)$ which represents the only existing subfield Z_1 of degree χ in Z_∞ has the discriminant $\mathfrak{p}^{(\chi-1)2}$. Consequently $\Omega\{t\}(x) \neq Z_1$, hence $Z = A_\infty(x)$ has degree χ over A_∞. The field A_∞ is not perfect, let L be its perfect closure. Then

---

[14] E. Witt, "Zyklische Körper und Algebren der Charakteristik p von Grade p^n," *Crelle*, vol. 176 (1937); H. L. Schmid, "Zur Arithmetik der zyklischen p-Körper," *Crelle*, vol. 176 (1937).

$$[ZL:L]=[Z:A_\infty]=\chi$$

according to Theorem 1.

Next we adjoin to L a transcendental element T such that either

α) $v(T)=$ irrational θ (for example an invariable element), or

β) $v(T)=$ rational, i. e. T is a pseudo limit of a pseudo convergent sequence of 2nd kind $\{a_i\}$ with elements in the algebraic closure of L.

In either case we know according to A. Ostrowski that the valuation $\mathfrak{p}_\infty$ of L is extended in a unique fashion to a valuation $\mathfrak{P}$ of $L(T)$ such that the field of residue classes of $L(T)$ with respect to $\mathfrak{P}$ is equal to Ω. For the value groups we have:

α) $\Gamma(L(T))=\{\Gamma(L),\Gamma^{(0)}\theta\}=\{\mathrm{P},\theta\Gamma^{(0)}\}$ and
β) $\Gamma(L(T))=\Gamma(L)=\mathrm{P}$, respectively.

Now let k be the $\mathfrak{P}$-adic closure of $L(T)$ then in either case $\mathfrak{o}/\mathfrak{P}=\Omega$ — $\mathfrak{o}$ the maximal order of all $\mathfrak{P}$-adic integers in k—and

α) $\Gamma(k)=\Gamma(L(T))=\{\mathrm{P},\theta\Gamma^{(0)}\}$ or
β) $\Gamma(k)=\Gamma(L(T))=\mathrm{P}$.

Since $L(T)$ is everywhere dense in k we can establish a one to one correspondence between the classes of normal simple algebras over $L(T)$ and k. We wish to prove that $\mathcal{G}^{(\chi)}(k)\neq 1$, hence it suffices to show the existence of proper χ-algebras over $L(T)$. We shall deal with the 2 cases separately.

Case α. The join $ZL(T)$ is a cyclic ramified extension of $L(T)$ and obviously $\Gamma(ZL(T))=\Gamma(L(T))$ for $v(x)\,\varepsilon\,\mathrm{P}$. Hence

$$\chi\Gamma(ZL(T))=\mathfrak{N}\Gamma(ZL(T))=\chi\Gamma(L(T))=\{\mathrm{P},\chi\theta\Gamma^{(0)}\}$$

where $\mathfrak{N}\Gamma(ZL(T))$ denotes the value group of the norms of all elements in $ZL(T)$. Consider now the algebra $(T,ZL(T)/L(T))$ with the generating relations

$$x^\chi-x=t^{-(1-\chi)},\ u^\chi=T,\ u^{-1}xu=x+1.$$

This algebra is not similar to $L(T)$ for then T has to be norm $\mathfrak{N}T'$ of a suitable element $T'\,\varepsilon\,ZL(T)$. This would imply for the values

$$\chi v(T)=\theta\,\varepsilon\,\mathfrak{N}\Gamma(ZL(T))=\{\mathrm{P},\theta\Gamma^{(0)}\},$$

such an inclusion cannot be true according to the ramification theory. Hence $(T,ZL(T)/L(T))\not\sim L(T)$, or

$$(T,Zk/k)\not\sim k,\text{ i. e. } G^{(\chi)}(k)\neq 1.$$

Case β. As before we consider the cyclic extension $ZL(T)$ of $L(T)$. But now we get

$$\chi\Gamma(ZL(T)) = \mathfrak{N}\Gamma(ZL(T)) = \Gamma(L(T)) = \Gamma(ZL(T)) = P.$$

The field $ZL(T)$ is ramified over $L(T)$ but we can no more measure the ramification by comparing the respective value groups.

Since it is sufficient to exhibit an example of a perfect field whose field of residue classes is algebraically closed and for which nevertheless $\mathfrak{g}^{(\chi)}(k) \neq 1$ we specialize $\chi = 2$ for sake of simplicity. Then $f(x) \equiv x^2 + x + t^{-3} = 0$. We consider the algebra $(T, ZL(T)/L(T))$ of degree 2 which is given by the generating relations

$$x^2 + x + t^{-3} = 0, \quad u^2 = T, \quad u^{-1}xu = x + 1.$$

An arbitrary element $a + bx$ of $ZL(T)$ has the characteristic equation

$$g(z) \equiv z^2 + az + a^2 + ab + b^2t^{-3} = 0,$$

hence $\mathfrak{N}(a + bx) = a^2 + ab + b^2t^{-3}$. The elements of $L(T)$ have the form $(\Sigma a_iT^i)(\Sigma c_iT^i)^{-1}$ where Σa_iT^i, Σc_iT^i are polynomials of T with coefficients in L. Suppose now that the element T is a norm, then we obtain a relation

$$(\Sigma a_iT^i)^2(\Sigma c_iT^i)^{-2} + (\Sigma a_iT_i)(\Sigma b_iT^i)(\Sigma c_iT^i)^{-2} \\ + (\Sigma b_iT_i)^2(\Sigma c_iT^i)^{-2}t^{-3} = T,$$

where Σc_iT^i denotes the common denominator of the elements a, b occurring in the equation $T = \mathfrak{N}(a + bx)$. Comparing the coefficients of the different powers of T one finds

$$t^3(a_0^2 + a_0b_0) + b_0^2 = 0$$
$$c_0b_1 + a_1b_0 = c_0, \cdots.$$

In order that T be a norm the diophantine equation $b_0^2 + b_0at^3 + a_0^2t^3 = 0$ must have a solution a_0, b_0 in L. We can transform this equation to a simple form by substituting

$$r = d + cb_0 \quad \text{where} \quad d, c \in L.$$

Then

$$rb_0 = db_0 + cb_0^2 = db_0 + b_0ca_0t^3 + ca_0^2t^3 \\ = ca_0^2t^3 + b_0(d + ca_0t^3),$$

hence r satisfies the equation

$$g(r) \equiv \begin{vmatrix} r + d & 0 \\ ca_0^2t^3 & r + d + ca_0t^3 \end{vmatrix}$$
$$= r^2 + d^2 + rca_0t^3 + dca_0t^3 + c^2a_0^2t^3$$
$$= r^2 + rca_0t^3 + (d^2 + dca_0t^3 + c^2a_0^2t^3) = 0.$$

Put $c = (a_0 t^3)^{-1}$, then

$$g(r) \equiv r^2 + r + (d^2 + d + t^{-3}) = 0.$$

The element $r = (a_0 t^3)^{-1} b_0$ satisfies the equation

$$r^2 + r + t^{-3} = 0$$

which is not solvable in L according to our general construction. But $T = \mathfrak{N}(a + bx)$ would imply that $a_0, b_0 \in L$, hence also $r \in L$ which cannot be true.

Thus

$$(T, ZL(T)/L(T)) \nsim L(T)$$

and

$$(T, Zk/k) \nsim k.$$

4. The structure of $\mathfrak{G}'(k)$. Whereas the structure of the group $\mathfrak{G}^{(\chi)}(k)$ is intimately related to the additive structure of the underlying field k the structure of the group $\mathfrak{G}(k)$ depends more upon the value group $\Gamma(k)$ and the multiplicative structure of the residue class field $\boldsymbol{k}$ as we shall see.

Let us assume that the field of residue classes $\boldsymbol{k}$ possesses for each integer n which is relatively prime to the characteristic χ of k a uniquely determined cyclic extension $\boldsymbol{U}_n$ of degree n moreover, we shall suppose that $\boldsymbol{k}$ is not real in the sense of Artin-Schreier if $\chi = 0$.[15] Let $\boldsymbol{f}_n(x) \equiv x^n + \boldsymbol{a}_1 x^{n-1} + \cdots + \boldsymbol{a}_n = 0$ be a generating equation of $\boldsymbol{U}_n$, then choose elements $a_i \in k$ such that $a_i \bmod \mathfrak{p} = \boldsymbol{a}_i$. The roots of the equation $f_n(x) \equiv x^n + a_1 x^{n-1} + \cdots + a_n = 0$ generate then a cyclic unramified extension U_n of degree n over k whose field of residue classes coincides with $\boldsymbol{U}_n$.[16] Since the fields $\boldsymbol{U}_n$ are uniquely determined the unramified extensions U_n whose degrees n are relatively prime to χ are uniquely determined in a fixed algebraically closed field of k.

Let U be the infinite unramified extension of k which contains the fields U_n. Then $\Gamma(U) = \Gamma(k)$ and the field of residue classes $\boldsymbol{U}$ possesses no algebraic extensions whose degrees are relatively prime to χ. We observe that the field U is not perfect, but it is relatively perfect provided that k is relatively perfect. This remark is obvious for any equation with coefficients in U can already be considered as an equation in a sufficiently large subfield U_n of U; since U_n is relatively perfect Hensel's criterion can be applied to the equation. Moreover, we have according to Theorem 2

---

[15] A field k shall be called quasi-algebraically closed if all proper and improper finite algebraic extensions of it are never centers of finite division algebras. This definition is weaker than the one usually given for quasi-algebraic closure.

[16] Cf. No. 7.

$$\mathfrak{G}(\bar{U}) = \mathfrak{G}(U) \times \bar{U}$$

where $\bar{U}$ denotes the perfect closure of U.

One shows as in the proof of Theorem 4 that the classes of algebras in $\mathfrak{G}'(k)$ which are split by U or $\bar{U}$ form a subgroup $\mathfrak{G}'_u(k)$.

THEOREM 5. *The class group of algebras* $\mathfrak{G}'_u(k)$ *which are split by unramified fields is isomorphic with the direct sum of* λ *additive groups which are isomorphic with the group of all rational numbers* mod 1 *whose denominators are relatively prime to* χ.

Proof. We verbally can repeat the proof of Theorem 4. Namely, according to the assumption concerning the unramified extensions we find that

$$\mathfrak{G}'_{U_n}(k) = \mathfrak{G}'_{U_{p_1}a_1}(k) \times \cdots \times \mathfrak{G}'_{U_{p_s}a_s}(k),$$

where $n = \prod_{i=1}^{s} p_i^{a_i}$ and $\mathfrak{G}'_{U_{p_i}}{}^{a_i}(k)$ denotes the group of all algebras which are split by $U_{p_i}{}^{a_i}$. Hence

$$\mathfrak{G}'_u(k) = \prod_{p\neq\chi}{}_x \mathfrak{G}'_{U(p)}(k)$$

where Π_x denotes the direct product of any finite number of algebras in the groups $\mathfrak{G}'_{U(p)}(k)$ consisting of all classes which can be represented by cyclic crossed products $(a, U_{p^i}/k)$. Consequently, it suffices to determine the structure of each $\mathfrak{G}'_{U(p)}(k)$. Since $U_{p^i} \subset U_{p^{i+1}} \subset \cdots \subset U_{(p)}$ we have to determine the structure of $\mathfrak{G}'_{U_{p^i}}(k)$. Let $(a, U_{p^i}/k)$ be an arbitrary algebra of the latter group. If $v(a) = \alpha = \beta + \sum_{i=1}^{\lambda} \gamma_i\theta_i$ then determine an element $b \varepsilon k$ such that $v(b) = \beta p^{-i}$ and fix elements $t_i \epsilon k$ such that $v(t_i) = \theta_i$. Hence

$$(a, U_{p^i}/k) \sim (b^{p^i}, U_{p^i}/k) \times (\prod_{i=1}^{\lambda} t_i^{\gamma_i}, U_{p^i}/k) \times (\epsilon, U_{p^i}/k),$$

the 1st and 3rd factors of the right side are similar to k as a repetition of the argument int he proof of Theorem 4 shows.

We only have to observe that algebras $(\epsilon, U_{p^i}/k)$ whose exponent equals the degree p^i do not exist. First suppose that $\chi \neq 0$, then $k(\epsilon^{1/p^i})$ has to coincide with U_{p^i} since there can be established a one to one correspondence between the unramified extensions of k and the cyclic extensions of $\boldsymbol{k}$. Hence $\epsilon = \mathfrak{N}\epsilon^{1/p^i}$ for -1 has to be a norm. Secondly, if $\chi = 0$ then the preceding argument applies for all primes $p \neq 2$. If $p = 2$ then the existence of proper division algebras over $\boldsymbol{k}$ is excluded by the assumption that $\boldsymbol{k}$ be not real in the sense of Artin-Schreier.

Consequently,

$$k^*/\mathfrak{N}U^*_{p^i} \cong \Gamma(k)/\mathfrak{N}\Gamma(U_{p^i}) = \Gamma(k)/p^i\Gamma(k)$$
$$\cong Z_1(p^i) \times \cdots \times Z_\lambda(p^i)$$

where $Z_j(p^i)$ denote cyclic groups of order p^i. Therefore $\mathfrak{G}'_{U(p)}(k)$ is isomorphic with the direct sum of λ additive groups of rational numbers mod 1 whose denominators are powers of p.

Remark. Combining Theorems 4 and 5 we see that the class groups of all algebras which are split by unramified fields is isomorphic with the direct sum of λ additive groups of all rational numbers mod 1.

LEMMA 4. *If the relatively perfect field k contains all n-th roots of unity—$(n, \chi) = 1$—then*

$$\mathfrak{G}'(\bar{U}) \cong \mathfrak{G}'(U) = \mathfrak{G}'(k) \times U \cong \mathfrak{G}'(k)/\mathfrak{G}'_u(k).$$

Proof. The proof of Lemma 2 yields that

$$k^*/k^{*n} \cong \Gamma/n\Gamma \times k^*/k^{*n} \quad \text{and} \quad U^*/U^{*n} \cong \Gamma/n\Gamma.$$

Consequently, the maximal abelian extension $K^{(n)}$ of U of exponent n is given by $K^{(n)} = U(t_1^{1/n}, \cdots, t_\lambda^{1/n}) = k(t_1^{1/n}, \cdots, t_\lambda^{1/n})U$. The ramification theory yields that any finite algebraic extension K whose degree is relatively prime to χ is contained in a suitable field $K^{(n)}$. Consequently, the class group $\mathfrak{G}'(U)$ is given by the set of subgroups $\mathfrak{G}'_{K^{(n)}}(U)$. Since the fields $K^{(n)}$ are extensions of fields over k we see that each algebra in $\mathfrak{G}'_{K^{(n)}}(U)$ can be obtained as the direct product of a suitable crossed product with the splitting field $K^{(n)}$ and of U. Hence $\mathfrak{G}'(U) = \mathfrak{G}'(k) \times U$.

LEMMA 5. *Let k be a relatively perfect field whose field of residue classes* **k** *possesses no algebraic extensions of degree n—$(n, \chi) = 1$—and whose value group $\Gamma(k)$ is either equal to*

$$\{\mathrm{P}, \mathrm{P}\chi^\infty\theta_1, \cdots, \mathrm{P}\chi^\infty\theta_\mu\} \quad \text{or} \quad \{\mathrm{P}, \mathrm{P}\chi^\infty\theta_1, \cdots, \mathrm{P}\chi^\infty\theta_\mu, \theta\Gamma^{(0)}\}$$

then

$$\mathfrak{G}'(k) = 1.$$

Proof. Suppose that $\Gamma(k) = \{\mathrm{P}, \mathrm{P}\chi^\infty\theta_1, \cdots, \mathrm{P}\chi^\infty\theta_\mu\}$ then k has no algebraic extension K_n of degree n—$(n, \chi) = 1$—whatsoever, for the Galois group of any such extension has to be isomorphic with the factor group

$$\Gamma(K)/\Gamma(k) = \Gamma(k)/\Gamma(k).$$

Consequently $G'(k) = 1$.

If $\Gamma(k) = \{P, P\chi^{\infty}\theta_1, \cdots, P\chi^{\infty}\theta_\mu, \theta\Gamma^{(0)}\}$ then the fields $K_n = k(t^{1/n})$ where $v(t) = \theta$ are the only existing extensions of degree n. Hence any algebra of $\mathfrak{B}'(k)$ must have a suitable field K_n as splitting field. Thus we have to consider the factor group $k^*/\mathfrak{N}K^*_n$. Obviously each unit ϵ is a n-th power of a suitable unit for the equation $x^n - \epsilon \bmod \mathfrak{p} = 0$ factors completely in $\boldsymbol{k}$. Also all elements a in k whose values $v(a)$ do not involve a multiple of θ are n-th powers: choose $b \in k$ such that $v(b) = v(a)n^{-1}$ then $a = b^n\epsilon = b^n\eta^n = (b\eta)^n$. Finally, there exists an element T in K_n whose norm is equal to t; namely choose $T = t^{1/n}$ if $n \not\equiv 0$ (2) or $T = t^{1/n}\omega$ with a suitable root of unity ω if $n \equiv 0$ (2). Hence $k^* = \mathfrak{N}K^*_n$, i. e. $\mathfrak{B}'(k) = 1$.

Let k be a relatively perfect field whose value group $\Gamma(k)$ is equal to $\{P, \theta_1\Gamma^{(0)}, \cdots, \theta_\lambda\Gamma^{(0)}\}$ and whose field of residue classes $\boldsymbol{k}$ is closed with regard to extensions of degree n, $(n, \chi) = 1$. We construct the following chain of infinite abelian extensions of k:

$$\begin{aligned} K_1 &= k(\langle t_2^{1/n}, \cdots, t_\lambda^{1/n}\rangle) \\ K_2 &= k(\langle t_3^{1/n}, \cdots, t_\lambda^{1/n}\rangle), \\ K_i &= k(\langle t_{i+1}^{1/n}, \cdots\ t_\lambda^{1/n}\rangle), \\ K_{\lambda-1} &= k(\langle t_\lambda^{1/n}\rangle) \end{aligned}$$

where n runs over all integers which are relatively prime to χ. We have $K_{i-1} = K_i(\langle t_i^{1/n}\rangle)$ and for the value groups

$$\begin{aligned} \Gamma(K_{i-1}) &= \{P, \theta_1\Gamma^{(0)}, \cdots, \theta_i\Gamma^{(0)}, \theta_{i+1}P\chi^{\infty}, \cdots, \theta_\lambda P\chi^{\infty}\} \\ &= \{\Gamma(K_i), \theta_i P\chi^{\infty}\}. \end{aligned}$$

All the fields K_i are relatively perfect.

THEOREM 6. *The group $\mathfrak{B}'(k)$ contains a series of subgroups*

$$\mathfrak{B}'(k) \supseteqq \mathfrak{B}'(k)_{K_1} \supset \cdots \supset \mathfrak{B}'(k)_{K_{i-1}} \supset \mathfrak{B}'(k)_{K_i} \supset \cdots \supset \mathfrak{B}'(k)_{K_{\lambda-1}}$$

such that

$$\begin{aligned} \mathfrak{B}'(k)_{K_{i-1}}/\mathfrak{B}'(k)_{K_i} &\cong \mathfrak{E}'(K_i)_{K_{i-1}} \\ &\cong (P\chi^{\infty} \bmod 1)^{(1)} + \cdots + (P\chi^{\infty} \bmod 1)^{(i-1)}. \end{aligned}$$

Proof. The group $\mathfrak{B}'(k)$ contains the subgroup $\mathfrak{B}'(k)_{K_i}$ which consists of all algebras which are split by K_i; therefore obviously $\mathfrak{B}'(k)_{K_{i-1}} \supset \mathfrak{B}'(k)_{K_i}$ according to the construction of the fields K_i. Evidently we have $(\mathfrak{B}'(k)_{K_{i-1}})_{K_i} = \mathfrak{B}'(k)_{K_i}$. Next we see that

$$\mathfrak{B}'(k) \times K_i = \mathfrak{B}'(K_i)$$

for $\boldsymbol{k}$ is closed with respect to algebraic extensions whose degrees are relatively prime to χ. Consequently,

$$\mathfrak{G}'(K_i) \cong \mathfrak{G}'(k)/\mathfrak{G}'(k)_{K_i}.$$

The group $\mathfrak{G}'(K_i)$ contains the group $\mathfrak{G}'(K_i)_{K_{i-1}}$ whose algebras are split by the fields $K_i(t_i^{1/n})$. Consequently the algebras

$$(t_j, K_i(t_i^{1/n})) = (t_j, k(t_i^{1/n})) \times K_i,$$

where $j = 1, \cdots, i-1$, furnish a complete set of generators of

$$\mathfrak{G}'(K_i)_{K_i(t_i^{1/n})} \subset \mathfrak{G}'(K_i)_{K_{i-1}}.$$

Next we have

$$\begin{aligned} \mathfrak{G}'(k)_{K_{i-1}} \times K_i = \mathfrak{G}'(K_i)_{K_{i-1}} &= \mathfrak{G}'(k)_{K_{i-1}}/(\mathfrak{G}'(k)_{K_{i-1}})_{K_i} \\ &\cong \mathfrak{G}'(k)_{K_{i-1}}/\mathfrak{G}'(k)_{K_i} \cong \mathfrak{G}'(K_i)_{K_{i-1}} \\ &\cong (\mathrm{P}\chi^{\infty} \bmod 1)^{(1)} + \cdots + (\mathrm{P}\chi^{\infty} \bmod 1)^{(i-1)}. \end{aligned}$$

The chain of subgroups has to start with $\mathfrak{G}'(k)_{K_1}$ according to Lemma 5.

LEMMA 6. *If* $(a_{\sigma,\tau}, K/k)$ *and* $(b_{\Sigma,\mathrm{T}}, N/k)$ *are two crossed products such that* $K \cap N = k$ *then* $(a_{\sigma,\tau}, K/k) \times (b_{\Sigma,\mathrm{T}}, N/k) \sim (c_{S,T}, KN/k)$ *where* $\| c_{S,T} \| = \| a_{\sigma,\tau} \| \otimes \| b_{\Sigma,\mathrm{T}} \|$ *is the Kronecker product of the two matrices* $\| a_{\sigma,\tau} \|$, $\| b_{\Sigma,\mathrm{T}} \|$ *which represent the factor sets.*

Proof. The algebra $(a_{\sigma,\tau}, K/k)$ has also the field KN as normal splitting field. Hence $(a_{\sigma,\tau}, K/k) \sim (a'_{S,T}, KN/k)$ where $a'_{S,T} = a'_{S',T}$, for $S \equiv S'(\{\Sigma\})$, $T \equiv T'(\{\Sigma\})$ and $a'_{\sigma\{\Sigma\},\tau\{\Sigma\}} = a_{\sigma,\tau}$ since the Galois group of KN/k is according to our assumption equal to the direct product of the Galois groups $\{\sigma\}$ and $\{\Sigma\}$ of K and N respectively.[17] The matrix $\| a'_{S,T} \|$ of degree $[KN:k]^2 = [K:k]^2[N:k]^2 = n^2m^2$ is equal to the Kronecker product of the matrix $\| a_{\sigma,\tau} \|$ and the unit matrix of degree $[N:k]^2$ as the relation $a'_{\sigma\{\Sigma\},\tau\{\Sigma\}} = a_{\sigma,\tau}$ implies. Hence $(a_{\sigma,\tau}, K/k) \sim (\| a_{\sigma,\tau} \| \otimes E_{m^2}, KN/k)$.

Similarly we obtain

$$(b_{\Sigma,\mathrm{T}}, N/k) \sim (E_{n^2} \otimes \| b_{\Sigma,\mathrm{T}} \|, KN/k)$$

when we take in account a fixed ordering of the elements of $\{\sigma\} \times \{\Sigma\}$ with regard to the direct factors. Combining the two representations we get

$$(a_{\sigma,\tau}, K/k) \times (b_{\Sigma,\mathrm{T}}, N/k) \sim (c_{S,T}, KN/k) = (\| a_{\sigma,\tau} \| \otimes \| b_{\Sigma,\mathrm{T}} \|, KN/k).$$

---

[17] D. Chap. V, § 4.

Let $\mathfrak{A}$ be an arbitrary simple algebra of the group $\mathfrak{G}'(k)$ whose degree is relatively prime to χ. Then every maximally commutative subfield K of $\mathfrak{A}$ is abelian. Let $K = Z_1 \cdots Z_\mu$ be an arbitrary but fixed representation of K as the join of relatively prime cyclic fields Z_i with the respective Galois groups $\{\sigma_i\}$. The algebra $\mathfrak{A}$ has the form

$$(a_{S,T}, K/k)$$

where $S, T, \cdots$ denote the elements of the Galois group $\{\sigma_1\} \times \cdots \times \{\sigma_\mu\}$ of K over k. We shall call components $a_{S,T}$ of the factor set *side components* if $S = \sigma_i^{s_i} \neq \epsilon$, $T = \sigma_j^{t_j} \neq \epsilon$ are not elements of the same subgroup $\{\sigma_i\}$.

We then can prove a general theorem which characterizes the classes of $\mathfrak{G}'(k)$ which are similar to direct products of classes which can be represented by cyclic algebras.

THEOREM 7. *A class of $\mathfrak{G}'(k)$ is similar to the direct product cyclic classes if and only if the given class possesses an abelian splitting field $K = Z_1, \cdots, Z_\mu$ such that the side components of a suitable associated factor set belonging to the class are equal to unity.*

Proof. Suppose that the algebra $\mathfrak{A}$ of the given class in $\mathfrak{G}'(k)$ possesses a splitting field K such that for a suitable representation of K as the join of cyclic fields Z_i there does exist a factor set $a_{S,T}$ whose side components (with respect to the fixed decomposition of K) are equal to unity. Then

$$u_{\sigma_i \nu_i} u_{\sigma_j \nu_j} = u_{\sigma_i{}^{\nu_i} \sigma_j{}^{\nu_j}} = u_{\sigma_j{}^{\nu_j} \sigma_i{}^{\nu_i}}$$

and

$$u^{-1}_{\sigma_i \nu_i} z^{(j)} u_{\sigma_i \nu_i} = z^{(j)}$$

for any pair (i, j) of different indices, $z^{(j)} \epsilon Z_j$. If $[K:k] = n = n_1 \cdots n_\mu = [Z_1:k] \cdots [Z_\mu:k]$, $[\mathfrak{A}:k] = n^2$ then all sums

$$\sum_{l=1}^{n_i} z_l{}^{(i)} u_{\sigma_i{}^l}$$

constitute a normal simple algebra $\mathfrak{A}^{(i)}$ of degree $n^{(i)}$, $i = 1, 2, \cdots, \mu$. The algebras $\mathfrak{A}^{(i)}$ are contained in $\mathfrak{A}$ and we can form the direct product $\mathfrak{A}^{(1)} \times \cdots \times \mathfrak{A}^{(\mu)}$ within $\mathfrak{A}$. Namely, the elements of different algebras $\mathfrak{A}^{(i)}$, $\mathfrak{A}^{(j)}$ permute with each other as a consequence of our assumption on the side components, and $[\mathfrak{A}:k] = \prod_{i=1}^{\mu} [\mathfrak{A}_i:k]$. Now we can normalize the factor set belonging to each component $\mathfrak{A}_i$ to the cyclic normal form

$$\mathfrak{A}_i = (a_i, Z^{(i}/k)$$

where $a_i \varepsilon k$. Since all units ϵ of k are n_i-th powers $(i = 1, \cdots, \mu)$ we can write

$$\mathfrak{A}_i = (\prod_{j=1}^{\lambda} t_j^{\nu_{ij}}, Z^{(i)}/k).$$

Moreover, it follows if we apply Lemma 5 that the original factor set $a_{S,T}$ is associated to a factor set which lies in the ground field k. However, $a_{S,T} \varepsilon k$ does not imply that A is similar to the direct product of cyclic algebras.

Conversely, if an algebra $\mathfrak{A}$ is similar to the direct product of cyclic algebras $(a_i, Z^{(i)}/k)$ when the splitting fields $Z^{(i)}$ are relatively prime over k $(i = 1, 2, \cdots, \mu)$ then $\mathfrak{A}$ is similar to an algebra which has the join $K = Z_1 \cdots Z_\mu$ as splitting field such that the side components of the corresponding factor set are equal to unity. Namely, each component $(a_i, Z^{(i)}/k)$ has the following matrix representing the complete factor set

$$\left\Vert \begin{array}{cccccccccc} 1 & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & 1 \\ \cdot & & & & & & & & \cdot & \\ \cdot & & & & & & & \cdot & & a_i \\ \cdot & & & & & & \cdot & & \cdot & \cdot \\ \cdot & & & & & \cdot & & \cdot & & \cdot \\ \cdot & & & & \cdot & & \cdot & & & \cdot \\ \cdot & & & \cdot & & \cdot & & & & \cdot \\ \cdot & & \cdot & & \cdot & & & & & \cdot \\ \cdot & \cdot & & \cdot & & & & & & \cdot \\ 1 & & a_i & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & a_i \end{array} \right\Vert = A_i$$

of degree $n_i = [Z_i : k]$.[18] Hence

$$\begin{aligned} \mathfrak{A} &\sim (a_1, Z^{(1)}/k) \times \cdots \times (a_\mu, Z^{(\mu)}/k) \\ &\sim (A_1 \otimes \cdots \otimes A_\mu, K/k) \end{aligned}$$

according to Lemma 6. The side components of $A_1 \otimes \cdots \otimes A_\mu$ obviously are equal to unity according to the structure of Kronecker products.

Let $\mathfrak{D}$ be an arbitrary normal division algebra over the relatively perfect field k, such that its degree n is relatively prime to the characteristic χ. Then the valuation $\mathfrak{p}$ of k can be extended to $\mathfrak{D}$ by defining the value $v(a) \varepsilon \Gamma(\mathfrak{D})$ of any element $a \neq 0$ in $\mathfrak{D}$ to be equal to

$$\frac{1}{n} v(\mathfrak{N}_{\text{red}}\; a)$$

where $\mathfrak{N}_{\text{red}}$ denotes the reduced norm of $\mathfrak{D}$ over k. Since $\Gamma(k)$ is assumed to have a finite irrational base $\theta_1, \cdots, \theta_\lambda$ we find that the factor group

---

[18] O. F. G. Schilling, "The structure of certain rational infinite algebras," *Duke Mathematical Journal*, vol. 3 (1937).

$$\Gamma(\mathfrak{D})/\Gamma(k)$$

is a finite abelian group whose invariants are proper or improper divisors of the degree n. Obviously, $[\Gamma(\mathfrak{D})/\Gamma(k) : 1] \leqq [\mathfrak{D} : k]$ for the representatives of the factor groups are values of elements which are linearly independent over k.

THEOREM 8. *A division normal algebra* $\mathfrak{D}$ *of degree* n *over* k *is equal to a cyclic algebra if and only if the factor group* $\Gamma(\mathfrak{D})/\Gamma(k)$ *contains at least one element of exact order* n.

Proof. Let a be an element of $\mathfrak{D}$ such that exactly $nv(a)$ lies in $\Gamma(k)$. Hence $v(a) = \alpha + \sum_{i=1}^{\lambda} \gamma_i \theta_i n^{-1}$ where at least one integer γ_i is relatively prime to n, $\alpha \varepsilon P$. Consequently, the field $k(a)$ is a maximally commutative subfield of $\mathfrak{D}$. It must be a cyclic extension, for

$$k(a) = k(b_1^{1/n_1}, \cdots, b_s^{1/n_s}), \qquad s > 1,$$

$b_i \varepsilon k$, $n_1 \cdots n_s = n$, $n_i /\!/ n$ would imply that $\Gamma(k(a))/\Gamma(k)$ has the invariants $(n_1, n_2, \cdots, n_s)$ whereas a is an element which has the exact order n. Therefore

$$\mathfrak{D} = (c, k(a)) = (c, k(b^{1/n}))$$

with a suitable quantity $c \neq 0$.

Conversely, if $\mathfrak{D} = (c, k(b^{1/n}))$ then

$$v(b^{1/n}) = \frac{1}{n} v(b) \, \varepsilon \, \Gamma(\mathfrak{D})$$

is an element of $\Gamma(\mathfrak{D})$ which has the exact order n with respect to $\Gamma(k)$.

THEOREM 9. *A crossed product* $(a_{\sigma,\tau}, K/k)$ *is never a division algebra if there exists a subset* a_{ρ_i,ρ_j} *of* $a_{\sigma,\tau}$ *which belongs to a cyclic subgroup* $\{\rho\}$ *of the Galois group* $\{\sigma, \tau, \cdots\}$ *and which has the form* $b_{\rho_i\rho_j} \varepsilon \rho_i, \rho_j$ *where* $v(b_{\rho_i,\rho_j}) \varepsilon \mathrm{P}$.

Proof. Let K_ρ be the subfield of K which belongs to the subgroup $\{\rho\}$ of $\{\sigma, \tau, \cdots\}$. Then

$$(a_{\sigma,\tau}, K/k) \times K_\rho \sim (a_{\rho_i,\rho_j}, K/K_\rho) = (\prod_{i=1}^{m} a_{\rho_i,\rho_j}, K/K_\rho)$$

if $[K : K_\rho] = m$.

The element $\prod_{i=1}^{m} a_{\rho^i,\rho} \varepsilon K_\rho$ has a rational value β according to assumption.

Choose now an element $b \varepsilon K_\rho$ whose value is equal to β/m, then

$$\prod_{i=1}^{m} a_{\rho^i,\rho} b^{-m}$$

is a unit ϵ_ρ of K_ρ and consequently

$$\begin{aligned}(\prod_{i=1}^{m} a_{\rho^i,\rho}, K_\rho/K_\rho) &\sim (b^m, K/K_\rho)(\epsilon_\rho, K/K_\rho)\\ &\sim (\epsilon_\rho, K/K_\rho)\\ &\sim K_\rho\end{aligned}$$

since each unit ϵ_ρ is a m-th power, or

$$(a_{\sigma,\tau}, K/k) \times K_\rho \sim K_\rho.$$

Suppose now that $(a_{\sigma,\tau}, K/k)$ would be a division algebra, then the degree of each of its splitting fields would be a multiple of $[K:k]$. We just showed that the field K_ρ of degree $[K:k]m^{-1} < [K:k]$ is a splitting field, but this contradicts our assumption that $(a_{\sigma,\tau}, K/k)$ were a division algebra.

We wish to conclude this section by some general remarks about the structure of normal simple algebras over a relatively perfect field k whose field of residue classes $\boldsymbol{k}$ is an arbitrary field of characteristic χ.

We have seen that there do exist normal simple algebras $\mathfrak{A}$ over k which are not split by any unramified extension U of k provided that $\lambda \geqq 2$. Thus we term a division algebra $\mathfrak{D}$ over k to be *totally ramified* if its algebra of residue classes $\boldsymbol{D}$ coincides with the field $\boldsymbol{k}$. A division algebra $\mathfrak{D}$ is called *unramified* if its algebra of residue classes $\boldsymbol{D}$ has the same degree over $\boldsymbol{k}$ as $\mathfrak{D}$ over k. It is a known fact that unramified algebras exist if and only if $\boldsymbol{k}$ is not quasi-algebraically closed.[19] As in the theory of fields which are perfect with respect to a discrete valuation we can prove the following existence theorem.[20]

THEOREM 10. *If $\boldsymbol{D}$ is a normal division algebra over $\boldsymbol{k}$ then there exists at least one unramified division algebra $\mathfrak{D}$ over k whose algebra of residue classes is equal to $\boldsymbol{D}$.*

THEOREM 11. *Each algebra of $\mathfrak{B}_u(k)$ is similar to the direct product of cyclic algebras and of an unramified division algebra.*

Proof. Let $\mathfrak{A} = (a_{\sigma,\tau}, U/k)$ be an arbitrary algebra of $\mathfrak{B}_u(k)$, then $v(a_{\sigma,\tau}) = \alpha(\sigma,\tau) + \sum_{i=1}^{\lambda} \gamma_i(\sigma,\tau)\mathfrak{e}_i$ where $\alpha(\sigma,\tau) \varepsilon \mathrm{P}$, $\gamma_i(\sigma,\tau) \varepsilon \Gamma^{(0)}$ since $\Gamma(U) = \Gamma(k)$. Determine elements $b_{\sigma,\tau} \varepsilon k$ such that

$$v(b_{\sigma,\tau}) = [U:k]^{-1}\alpha(\sigma,\tau)$$

---

[19] See R. Brauer, "Über die Konstruktion der Schiefkörper, die von endlichen Rang in Bezug auf ein gegebenes Zentrum sind," § 7, Theorem 12, *Crelle*, vol. 168 (1932).

[20] See T. Nakayama, "Divisionsalgebren über diskret bewerkten perfekten Körpern," *Crelle*, vol. 177 (1937).

and fix elements t_i such that $v(t_i) = \theta_i$. Then

$$a_{\sigma,\tau} b_{\sigma,\tau} \prod_{i=1}^{\lambda} t_i^{-\gamma_i(\sigma,\tau)}$$

is a factor set consisting of units $\epsilon_{\sigma,\tau} \varepsilon U$ for $b_{\sigma,\tau} \varepsilon k$ and

$$\gamma_i(\sigma, \tau\rho) + \gamma_i(\tau, \rho) = \gamma_i(\sigma\tau, \rho) + \gamma_i(\sigma, \tau) \qquad (i = 1, 2, \cdots, \lambda).$$

Hence

$$\begin{aligned} \mathfrak{A} &\sim (\epsilon_{\sigma,\tau}, U/k)(b_{\sigma,\tau}^{-1}, U/k)(\prod_{i=1}^{\lambda} t_i^{\gamma_i(\sigma,\tau)}, U/k) \\ &\sim (t_1^{\gamma_1(\sigma,\tau)}, U/k) \times \cdots \times (t_\lambda^{\gamma_\lambda(\sigma,\tau)}, U/k)(\epsilon_{\sigma,\tau}, U/k) \\ &\sim (t_1, Z_1/k) \times \cdots \times (t_\lambda, Z_\lambda/k)(\epsilon_{\sigma,\tau}, U/k) \end{aligned}$$

where Z_i denote the cyclic subfields of U which are induced by the characters related to the sets of addends $\gamma_i(\sigma, \tau)$. The algebras $(t_i, Z_i/k)$ whose generating automorphisms are normalized are division algebras whose algebras of residue classes are equal to $\mathbf{Z}_i$. The algebra $(\epsilon_{\sigma,\tau}, U/k)$ is obviously similar to an unramified division algebra.

If we suppose that all unramified are products of cyclic algebras—which amounts to assuming that all algebras over $\boldsymbol{k}$ are direct products of cyclic algebras—then each algebra of $\mathfrak{G}_u(k)$ is representable as the direct product of cyclic algebras. In general, however, $\mathfrak{G}_u(k)$ need not exhaust the group of all classes of algebras which can be represented as direct products of cyclic algebras. Namely, if $\lambda = 4$ then there exists a division algebra of degree 4 which is totally ramified:

$$(t_1, k(\sqrt{t_2})/k)(t_3, k(\sqrt{t_4})/k).$$

All classes which can be represented by unramified division algebras evidently form a group. The classes which can be represented by completely ramified division algebras do not form a group. Consider the following example. Assume that k contains the n-th roots of unity and that there does exist a unit ϵ in k which is not the norm of an element of the cyclic extension $k(t_2^{1/n})$ $(\lambda = 2)$. The two algebras

$$\mathfrak{A}_1 = (t_2, k(t_1^{1/n})/k) \quad \text{and} \quad \mathfrak{A}_2 = (t_2, k((\epsilon t_1)^{1/n})/k)$$

obviously are division algebras. They are similar to the algebras

$$(t_1^{-1}, k(t_2^{1/n})/k) \quad \text{and} \quad (t_1^{-1}\epsilon^{-1}, k(t_2^{1/n})/k)$$

respectively. Then

$$\mathfrak{A}_1^{-1} \times \mathfrak{A}_2 \sim (\epsilon^{-1}, k(t_2^{1/n})/k) \sim (t_2, k(\epsilon^{1/n})/k),$$

the resulting algebra is a division algebra whose residue algebra is equal to the field of residue classes belonging to the unramified cyclic extension $k(\epsilon^{1/n})$ of k.

It is easy to see that in case of $\lambda = 0$ there do exist only unramified algebras in the class group $\mathfrak{G}'(k)$. Then each algebra $\mathfrak{A}$ in $\mathfrak{G}'(k)$ has an unramified normal splitting field U,

$$\mathfrak{A} = (a_{\sigma,\tau}, U/k)$$

and

$$\mathfrak{A} \sim (\epsilon_{\sigma,\tau}, U/k)$$

where

$$\epsilon_{\sigma,\tau} = a_{\sigma,\tau} b_{\sigma,\tau}$$

when

$$v(b_{\sigma,\tau}) = [U:k]^{-1} v(a_{\sigma,\tau}).$$

It is also possible to develop the arithmetic theory of simple algebras $\mathfrak{A} = \mathfrak{D}_r$. However, in the proofs one has to avoid any argument which may necessitate the use of bases of orders and ideals. The existence of at least one maximal order $\mathfrak{M}$ which contains a given order of highest rank in $\mathfrak{A}$ can be proved by applying well ordering. Since $\Gamma(k)$ is everywhere dense in the additive group of all real numbers it can be shown that an algebra $\mathfrak{A} = \mathfrak{D}_r$ $(r > 1)$ contains continuously many maximal orders. We do not wish to develop the arithmetic theory here for it has only indirect bearing upon the problem of local class field theory.

5. Generalization of classical local class field theory. We showed in section 3 by two examples that there exist under certain circumstances proper division algebras in $\mathfrak{G}'(k)$ if the field of residue classes $\boldsymbol{k}$ is algebraically closed, although $\lambda = 0, 1$. We can expect for this reason a satisfactory and complete characterization of the abelian fields K over k whose degrees are relatively prime to the characteristic χ. Theorems 5, 6, and 7 of the last section show that the number λ of rationally independent irrational generators of the value group $\Gamma(k)$—they may be ordered according to increasing absolute values—can assume the following determinations in order that the group $\mathfrak{G}'(k)$ be isomorphic with the additive group $\mathrm{P}_{\chi}{}^{\infty}$ mod 1:

Case i). For each n there exists a uniquely determined extension $\boldsymbol{U}_n$ of degree n over $\boldsymbol{k}$ then $\lambda = 1$, and

Case ii). The field of residue classes $\boldsymbol{k}$ is algebraically closed then $\lambda = 2$.

We shall put more stress on the first case since one readily can extend the methods used there to the second case. First of all, Lemma 5 shows that the group $\mathfrak{G}'(U)$ which belongs to the unramified closure of the given ground fields is equal to one. Since $\mathfrak{G}'(k) \times U \subseteq \mathfrak{G}'(U)$ we see that all classes of algebras in $\mathfrak{G}'(k)$ possess unramified cyclic splitting fields. Therefore $\mathfrak{G}'(k) = \mathfrak{G}'_u(k)$ and hence

$$\mathfrak{G}'(k) \cong \mathrm{P}_\chi{}^\infty \bmod 1$$

according to Theorem 5. Thus we proved

THEOREM 12. *The class group of algebras* $\mathfrak{G}'(k)$ *over a relatively perfect field whose value group* $\Gamma(k) = \{\mathrm{P}, \theta_1\Gamma^{(0)}\}$ *and which possesses for each integer* n —$(n, \chi) = 1$ —*a unique unramified extension of degree* n, *is isomorphic to the additive group* $\mathrm{P}_\chi{}^\infty \bmod 1$.

In the following statements it shall be understood that the relative degree of any field K and the degree of any division algebra considered, shall be relatively prime to the characteristic χ.

THEOREM 13. *The Galois group of any abelian extension* K *over* k *is isomorphic with the norm class group* $k^*/\mathfrak{N}K^*$.

Proof. We first show that each field K of degree n is splitting field of each normal division algebra $\mathfrak{D}$ of degree n. Let $\mathfrak{D}$ be an arbitrary division algebra of degree n then according to Theorem 12

$$\mathfrak{D} = (t^\nu, U_n/k)$$

where $(\nu, n) = 1$ and $v(t) = \theta$, for a fixed choice of the generating automorphism of the Galois group of U_n with respect to k.

The general ramification theory of A. Ostrowski yields in the particular case we have to consider, that the inertial field $K(I)$ of the given field K is a field U_f, where f/n. Moreover, $[K:K(I)] = n/f = e$ and hence as a consequence of our assumption on $\Gamma(k)$

$$t = T^e\mathrm{E}$$

where T denotes an element of K such that $v(\mathfrak{N}T) = \theta$, E a suitable unit of K.

Consider now the algebra [21]

$$\begin{aligned}
\mathfrak{D} \times K &\sim (T^e\mathrm{E}, U_nK/K) \\
&\sim (T^e, U_nK/K)(\mathrm{E}, U_nK/K) \\
&\sim (T, U_nK/K)^e(\mathrm{E}, U_nK/K) \\
&\sim (\mathrm{E}, U_nK/K) \\
&\sim K
\end{aligned}$$

---

[21] Compare with C. Chevalley, "La théorie du symbole de restes normiques," *Crelle*, vol. 169 (1933).

where $(T^e, U_nK/K) \sim K$ as a consequence of $[U_nK : K] = e$ and where $(\mathrm{E}, U_nK/K) \sim K$ since the fields of residue classes $\boldsymbol{k}$ and a fortiori $\boldsymbol{K}$ are quasi-algebraically closed. Now let Z_n be an arbitrary cyclic extension of degree n then

$$G'(k)_{Z_n} = G'(k)_{U_n} \cong \mathrm{Z}(n)$$

for

$$\mathfrak{D} \times Z_n \sim Z_n$$

where $\mathfrak{D}$ is any generator of the cyclic class group $\mathfrak{G}'(k)_{U_n} \subset \mathrm{P}_\chi^\infty \bmod 1$; $\mathrm{Z}(n)$ denotes a cyclic group of order n. Consequently,

$$k^*/\mathfrak{N}Z^*_n \cong \mathrm{Z}(n)$$

holds for any cyclic field Z_n. Moreover, the same result can be proved for the cyclic fields over arbitrary finite extensions L of k since the residue fields $\boldsymbol{L}$ are still quasi-algebraically closed and possess for each n exactly one unramified extension of degree n.

Next let K be an arbitrary normal extension of degree n over k. The ramification theory yields immediately that the Galois group of K over k is solvable. Using a composition chain of the Galois group such that the respective factor groups are cyclic we find that we can find a chain of subfields $k \subset \cdots \subset K_{i-1} \subset K_i \subset \cdots \subset K$ between k and K such that the fields K_i are cyclic extensions of the fields K_{i-1}. Application of the previously obtained result concerning the norm factor groups of cyclic extensions yields

$$K^*_{i-1}/\mathfrak{N}_i K^*_i \cong \mathrm{Z}([K_i : K_{i-1}])$$

when $\mathfrak{N}_i$ denotes the relative norm taken from K_i to K_{i-1}. Consequently

$$[k^* : \mathfrak{N}K^*] \leqq [K : k]$$

since the norms are transitive multiplicative functions. Now let K be an abelian extension of degree n with the Galois group $\Sigma = \{\sigma, \tau, \cdots\}$. Then there exists at least one factor set $a_{\sigma,\tau}$ of elements in K such that the n-th and no earlier power of it is equivalent to the unit factor set. The existence of $a_{\sigma,\tau}$ is the immediate consequence of the similarity relation

$$\mathfrak{D} \times K \sim K$$

where $\mathfrak{D}$ denotes an arbitrary generating algebra of $\mathfrak{G}'(k)_{U_n}$. We form the elements

$$b_\sigma = \prod_{\tau \epsilon \Sigma} a_{\sigma,\tau},$$

they determine an isomorphism between Σ and a subgroup of $k^*/\mathfrak{N}K^*$.[22] Since $[k^* : \mathfrak{N}K^*] \leqq n$ we have

$$\Sigma \cong k^*/\mathfrak{N}K^*$$

Thus we proved the isomorphism theorem for the abelian extensions of k.

Remark. The group of *all* algebras which have unramified splitting fields, i. e. $\mathfrak{B}'_u(k)$ is isomorphic with the additive group of all rational fractions mod 1 provided that also for the degrees χ^m there exist uniquely determined cyclic unramified fields.

In the classical theory the isomorphism $\Sigma \cong k^*/\mathfrak{N}K^*$ is given by a definite realization by the norm residue symbol.[23] The possibility of normalizing the aforementioned isomorphism depends upon the fact that an arithmetically significant automorphism which generates the cyclic Galois group belonging to the unramified fields can be found. Thus, if we make no special assumption on the algebraic structure of the residue class field **k** we have to find a substitute for the so-called Frobenius automorphism.

Let U be the infinite cyclic unramified extension of k which is equal to the join of all unramified extensions U_n of k. The Galois group of U with regard to k is then a totally disconnected compact group $O = \{o, \cdots\}$.[24] It is easy to give an explicit representation of the group O. Let then

$$k \subset U^{(1)} \subset U^{(2)} \subset \cdots \subset U^{(i-1} \subset U^{(i)} \subset \cdots \subset U$$

be an arbitrary—but necessarily enumerable—tower of unramified fields approximating U and let $O_i = \{o_i, \cdots\}$ be the Galois groups of $U^{(i)}$ with respect to k. Since $U^{(i)}/U^{(i-1)}$ and $U^{(i)}/k$ are normal the substitutions o_i of O_i induce exactly the substitutions o_{i-1} of O_{i-1}, or O_{i-1} is a homomorphic map $O_i * \phi_i$ of O_i. Then O can be isomorphically represented by the sequences $[o_1, o_2, \cdots, o_{i-1}, o_i, \cdots]$ of elements $o_i \varepsilon O_i$ such that $o_{i-1} = o_i * \phi_i$.

In order to obtain a substitute for the Frobenius substitution let us fix an arbitrary approximation of U and an arbitrary generating automorphism o_1 of U_1. Then we take an arbitrary but fixed element ψ of our uniquely determined representation of O which induces the substitution o_1 in U_1. Since each unramified field U_n is contained in U the automorphism ψ_n determines uniquely a *generating* substitution ψ_n of the Galois group belonging to U_n over k. We shall use this element ψ_n in order to normalize the representation of the elements of the class group $G'(k)_{U_n}$. According to the theory of cyclic crossed products we know that any cyclic algebra (a, U_n, σ_n) is equal to (a^r, U_n, ψ_n) where $\sigma_n = \psi_n{}^s$ is a generating substitution of the Galois group and $sr \equiv 1 \pmod{n}$. Hence each algebra of $G'(k)_{U_n}$ is similar to a power of the normalized division algebra (t, U_n, ψ_n) where t denotes a fixed element of k whose value is equal to $\theta \,\varepsilon\, \Gamma(k)$. It is readily seen that this normalization for a fixed degree n is consistent with the normalization of all other degrees m which are multiples of n; one only has to take in account the familiar rules concerning cyclic crossed products. The substitutions ψ_n and normalizations (t, U_n, ψ_n) have *algebraically* the same significance as the classical normalizations with respect to the Frobenius substitution. Consequently, we are able to define a generalized *norm residue symbol.* Let (a, Z_n, S_n) be an algebra which belongs to the class group of algebras split by the arbitrary but fixed cyclic field Z_n; S_n a generating automorphism of the Galois group. Suppose now that

---

[22] Yasuo Akizuki, "Eine homomorphe Zuordnung der Elemente der galoisschen Gruppe zu den Elementen einer Untergruppe der Normenklassengruppe," *Mathematische Annalen*, vol. 112 (1936).

[23] Cf. paper quoted under No. 19.

[24] J. Herbrand, "Théorie arithmétique des corps de nombres de degré infini, II. Extensions algébriques de degré infini," *Mathematische Annalen*, vol. 108 (1933).

$$(a, Z_n, S_n) \sim (t, U_n, \psi_n)^{\alpha},$$

then we define the norm residue symbol (a, Z_n) to be the substitution $S_n{}^{\alpha}$. Using this definition we verbally can take over the theorems of the classical local class field theory. One also observes going over the known proof of the existence, ordering and translation theorems that the latter are preserved for the type of relatively perfect fields we consider under case one. Thus we can state

THEOREM 14. *Let k be a relatively perfect field whose value group $\Gamma(k)$ is isomorphic with $\{\mathrm{P}, \theta\Gamma^{(0)}\}$ and whose residue class field k admits for each n exactly one cyclic extension of degree n, $(n, \chi) = 1$. Then the law of reciprocity, existence, ordering and translation theorems of the classical local class field theory hold for k as ground field.*

In case two where the field of residue classes $\boldsymbol{k}$ is algebraically closed and where $\lambda = 2$ we can repeat all preceding considerations. However, it is necessary to consider either $k(\langle t_1{}^{1/n}\rangle)$ or $k(\langle t_2{}^{1/n}\rangle)$ as the universal splitting field of the group $\mathfrak{G}'(k)$. In either case we obtain the equivalent of the local class field theory. It only has to be pointed out that the generalization of the Frobenius automorphism has no more an invariantive arithmetical significance.

Remark. It is possible to establish the results of classical local class field theory even if the assumptions made on page 695 are slightly weakened. Namely, we can generalize the value group Γ to arbitrary subgroups of the additive group of all real numbers with the restriction that

$$\Gamma/n\Gamma \cong \frac{1}{n}\Gamma/\Gamma \cong Z(n) \text{ for all } n \text{ in case i)},$$

and

$$\Gamma/n\Gamma \cong \frac{1}{n}\Gamma/\Gamma \cong Z(n) \times Z(n) \text{ for all } n \text{ in case ii)}.$$

However, it should be observed that the representatives of the factor groups can well be different for different values of n. Nevertheless, for powers of a single prime one can choose one and the same representatives, if the integer n is composite then the representatives can be built up additively from the representatives of the prime factors of n. In generalizing the proofs of the preceding theorems this fact has to be taken in account, in particular with regard to the selection of elements of the perfect fields whose values are equal to representatives of certain factor groups. Going over the proofs of the preceding theorems one easily observes that the generalization is obvious.

6. Examples of perfect fields with prescribed value group. We wish to give in this section several examples of perfect fields k whose value groups $\Gamma(k)$ are isomorphic to a prescribed group $\{\mathrm{P}, \theta_1\Gamma^{(0)}, \cdots, \theta_\lambda\Gamma^{(0)}\}$ of real numbers and whose fields of residue classes $\boldsymbol{k}$ are given too.

Let $\boldsymbol{k}$ be an arbitrary field of characteristic χ. Consider then the field of rational functions $\boldsymbol{k}(x_0, x_1, \cdots, x_\lambda)$ and put

$$(*) \qquad \begin{aligned} v(x_0) &= 1 \\ v(x_i) &= \theta_i \qquad (i = 1, \cdots, \lambda) \end{aligned}$$

where θ_i are given real numbers which are rationally independent. The set of $\lambda + 1$ relations (*) defines then a valuation $\mathfrak{p}$ on $\boldsymbol{k}(x_0, x_1, \cdots, x_\lambda)$ whose valuation ring contains the ring of polynomials $\boldsymbol{k}[x_0, x_1, \cdots, x_\lambda]$ and whose field of residue classes obviously coincides with the given field $\boldsymbol{k}$. Next we adjoin to $\boldsymbol{k}(x_0, x_1, \cdots, x_\lambda)$ the radicals x_0^{1/p^j} where p runs over all primes which are different from the characteristic χ and where $j = 1, 2, \cdots$. The resulting field $\boldsymbol{k}(x_0, x_1, \cdots, x_\lambda)(\langle x_0^{1/p^j}\rangle)$ admits then, according to the general valuation theory, a valuation whose value group is equal to

$$\{\mathrm{P}_{\chi^\infty}, \theta_1\Gamma^{(0)}, \cdots, \theta_\lambda\Gamma^{(0)}\}$$

and its field of residue classes coincides with $\boldsymbol{k}$.

Now we distinguish two cases

(i) $\chi = 0$ (ii) $\chi \neq 0$

Case i). We adjoined all n-th roots of x_0 and thus obtained an infinite algebraic extension K of $\boldsymbol{k}(x_0, x_1, \cdots, x_\lambda)(\langle x_0^{1/n}\rangle)$ on which there is defined a valuation $\mathfrak{P}$ whose value group is equal to $\{\mathrm{P}; \theta_1\Gamma^{(0)}, \cdots, \theta_\lambda\Gamma^{(0)}\}$ and whose field of residue classes is still equal to $\boldsymbol{k}$. Taking the $\mathfrak{P}$-adic closure of K we obtain a perfect field k which has the desired properties.

Case ii). We adjoin to $\boldsymbol{k}(x_0, x_1, \cdots, x_\lambda)(\langle x_0^{1/p^j}\rangle)$ the roots of the cyclic equations

$$y_\nu^\chi - y_\nu = a_\nu \text{ where } a_\nu = (1/x_0, \cdots, 1/x_0) \text{ is a vector of } \nu \text{ elements,} \quad (\nu = 1, 2, \cdots).$$

The valuation $\mathfrak{p}$ of $\boldsymbol{k}(x_0, x_1, \cdots, x_\lambda)$ possesses again at least one extension $\mathfrak{P}$ in the infinite algebraic extension

$$K = \boldsymbol{k}(x_0, x_1, \cdots, x_\lambda)(\langle x_0^{1/p^j}\rangle)(\langle y_\nu\rangle).$$

The $\mathfrak{P}$-adic closure k of K is then a field which possesses the required value group and residue field.

It is also possible to prove the existence of fields k with given value group by an entirely different method. Let now $\boldsymbol{k}$ be an algebraically closed field of characteristic χ. Consider the field

$$\boldsymbol{k}(x_0, y; x_1, \cdots, x_\lambda)$$

where

$$y = \sum_{i=0}^{\infty} \boldsymbol{a}_i x_0^{r_i}$$

and where $\{r_i\} \to \infty$ is a sequence of rational numbers whose denominators are powers of primes $p \neq \chi$, so that ultimately arbitrarily high powers of all the primes p occur in the denominators. Obviously, infinitely many such selections can be made for the sequence $\{r_i\}$. Put then

$$\begin{aligned} v(x_0) &= 1 \\ v(y) &= r_0 \\ v(x_i) &= \theta_i \qquad (i = 1, \cdots, \lambda). \end{aligned}$$

We thus determine a valuation of the field $L = \boldsymbol{k}(x_0, y; x_1, \cdots, x_\lambda)$ whose field of residue classes is equal to $\boldsymbol{k}$. The value group belonging to $\mathfrak{p}$ contains the components

$$\Gamma^{(0)} \leqq r_0\Gamma^{(0)}, \theta_1\Gamma^{(0)}, \cdots, \theta_\lambda\Gamma^{(0)}.$$

But we can prove more:

$$\Gamma(L) = \{\chi^{\infty}\mathrm{P}; \theta_1\Gamma^{(0)}, \cdots, \theta_\lambda\Gamma^{(0)}\}.$$

In order to see this we only have to prove the existence of elements in L whose values are equal to $1/p^j$ for each $p \neq \chi$ and $j = 1, 2, \cdots$. We apply for this purpose an argument which is used in the theory of Puiseux developments of algebraic function fields of two variables. Namely, consider the subfield $\boldsymbol{k}(x_0, y)$ of two variables. It is contained in the perfect field M of all formal power series $\sum_{i=0}^{\infty} \boldsymbol{b}_i x_0^{s_i}$ where $\boldsymbol{b}_i \in \boldsymbol{k}$ and $\{s_i\} \rightarrow \infty$ are arbitrary sequences of rational numbers. The perfectness of this field can readily be proved. The element $y = \Sigma \boldsymbol{a}_i x_0^{r_i}$ which is transcendental over $\boldsymbol{k}(x_0)$ is contained in the field M. We have to show that suitable rational functions of x_0 and y have the values $1/p^j$. Let then $\boldsymbol{a}_i x^{q_i/p^j}$ be *the* term in $y = \Sigma \boldsymbol{a}_i x_0^{r_i}$ whose denominator is exactly equal to p^j, according to our assumption on the sequence $\{r_i\}$ such a term is uniquely determined. We next suppose, in order to apply induction with regard to i, that we already proved the existence of rational functions of x_0, y whose values are equal to $1/p^l$ where p and l belong to the exponents r_ν, $\nu < j$. We write the series $y = \Sigma \boldsymbol{a}_i x_0^{r_i}$ in the form

$$\begin{aligned} y &= \sum_{\mu=0}^{i-2} \boldsymbol{a}_\mu x_0^{r_\mu} + \boldsymbol{a}_{i-1} x_0^{r_{i-1}} + \boldsymbol{a}_i x_0^{r_i} + \sum_{\mu=i+1}^{\infty} \boldsymbol{a}_\mu x_0^{r_\mu} \\ &= S' + \boldsymbol{a}_{i-1} x_0^{r_{i-1}} + \boldsymbol{a}_i x_0^{r_i} + S''. \end{aligned}$$

Next we consider the element

$$\bar{y} = S' + \boldsymbol{a}_{i-1} x_0^{r_{i-1}}$$

and its conjugates

$$\bar{y}_\rho = S'^{\sigma_\rho} + \boldsymbol{a}_{i-1} (x_0^{r_{i-1}})^{\sigma_\rho}$$

where σ_ρ denotes the $q_{i-1}^{l_{i-1}} = \tau$ automorphisms of the field M given by $x_0^{1/\tau} \rightarrow \zeta_\tau x_0^{1/\tau}$, ζ_τ denoting a fixed τ-th root of unity. $(r_{i-1} = r'_{i-1}/\tau)$ form the product

$$\prod_{\rho=1}^{\tau} (y - \bar{y}_\rho).$$

As in the classical theory of Puiseux developments it is seen that $\Pi(y - \bar{y}_\rho)$ is an element of $\boldsymbol{k}(x_0, y)$. The first term of the product $\Pi(y - \bar{y}_\rho)$ has then a first term $x_0^{s/m}$ where the denominator m is exactly divisible by p^j. There still may occur powers of other primes in m, however, they belong to the rational numbers r_μ $(\mu = 0, \cdots, i-2)$. We write then s/m as a sum of elementary fractions

$$\sum_{\mu=0}^{i-2} s_\mu p_\mu^{-l_\mu} + s' p^{-l}$$

where p_μ are the denominators of r_μ. According to induction, there exist elements z_μ in $\boldsymbol{k}(x_0, y)$ such that

$$v(z_\mu) = 1/p_\mu^{l_\mu}.$$

The element

$$\prod_{\mu=0}^{i-2} z_\mu^{-s_\mu} \prod (y - \bar{y}_\rho) = z$$

has the value

$$s'/p^l, \text{ or } z = \boldsymbol{c} x_0^{s'/p^l} + S'''$$

where $\boldsymbol{c} \neq 0$ and S''' consists of powers of x_0 whose exponents are greater than s'/p^l and whose denominators satisfy the same conditions as those of S''. Let s'' be a solution of the congruence

$$s's'' \equiv 1 \ (\mathrm{mod}\ p^l), \text{ i.e. } s's'' = 1 + gp^l, \ g \text{ an integer.}$$

The element $z^{s''}$ has the form

$$\boldsymbol{c}^{s''} x_0^{1/p^l} + g + S^{(4)},$$

consequently the element $z^{s''} x_0^{-g}$ has the value $1/p^l$. Thus we see by repeating this process that any power of a prime $p \neq \chi$ can occur in the denominators of the values belonging to suitable *rational* functions of x_0 and y. Therefore,

$$\Gamma(\boldsymbol{k}(x_0, y)) = \{\mathrm{P}\chi^\infty\}$$

and hence

$$\Gamma(\boldsymbol{k}(x_0, y; x_1, \cdots, x_\lambda)) = \{\mathrm{P}\chi^\infty, \theta_1\Gamma^{(0)}, \cdots, \theta_\lambda\Gamma^{(0)}\}.$$

Again as before we have to distinguish two cases according to the value of the characteristic χ. If $\chi = 0$ then we find that already

$$\Gamma(\boldsymbol{k}(x_0, y; x_1, \cdots, x_\lambda)) = \{\mathrm{P}, \theta_1\Gamma^{(0)}, \cdots, \theta_\lambda\Gamma^{(0)}\}.$$

Consequently, the perfect closure k of $\boldsymbol{k}(x_0, y; x_1, \cdots, x_\lambda)$ has the desired properties. We remark that the perfect closure contains also all formal power series $\Sigma \boldsymbol{b}_i x_0^{R_i}$ where $\boldsymbol{b}_i \in \boldsymbol{k}$ and $R_i \to \infty$ denotes any sequence of rational numbers whose denominators are relatively prime to χ. If $\chi \neq 0$ then we form the infinite algebraic extension N of $\boldsymbol{k}(x_0, y; x_1, \cdots, x_\lambda)$ which contains all solutions of the equations

$$y_\nu^\chi - y_\nu = a_\nu, \qquad (\nu = 1, 2, \cdots).$$

The valuation $\mathfrak{p}$ of $\boldsymbol{k}(x_0, y; x_1, \cdots, x_\lambda)$ has then an extension $\mathfrak{P}$ in N whose value group is equal to $\{\mathrm{P}, \theta_1\Gamma^{(0)}, \cdots, \theta_\lambda\Gamma^{(0)}\}$ and whose residue class field is equal to $\boldsymbol{k}$.

The last construction can also be applied if the field of residue classes $\boldsymbol{k}$ is not algebraically closed. We start again with a series $y = \Sigma \boldsymbol{a}_i x_0^{r_i}$, then we extend $\boldsymbol{k}(x_0, y)$ to its unramified algebraic closure $\bar{\boldsymbol{k}}(x_0, y)$ and apply the previous considerations to $\boldsymbol{k}(x_0, y)$. We find that $\Gamma(\bar{\boldsymbol{k}}(x_0, y)) = \{\chi^\infty \mathrm{P}\}$. Since $\bar{\boldsymbol{k}}(x_0, y)$ is an unramified extension of $\boldsymbol{k}(x_0, y)$ with respect to the valuation as given by $v(x_0) = 1$, $v(y) = r_0$ we have

$$\Gamma(\boldsymbol{k}(x_0, y)) = \Gamma(\bar{\boldsymbol{k}}(x_0, y)), \text{ i.e.}$$

there exist rational functions of x_0, y whose values are equal to $1/p^l$ where $(p, \chi) = 1$.

In order to construct perfect fields of characteristic 0 with prescribed value group where fields of residue classes have characteristic χ, we start with a χ-adic number field k and adjoin transcendental quantities $x_1, \cdots, x_\lambda$ to it. It is evident how the construction can be completed. Also it is fairly obvious how to construct perfect fields k whose value groups $\Gamma(k)$ have the type

$$\{G_0^{-1}\Gamma^{(0)}, G_1^{-1}\Gamma^{(0)}\theta_1, \cdots \quad G_\lambda^{-1}\Gamma^{(0)}\theta_\lambda\}$$

where the G_i are Steinitz-numbers describing the possible denominators of the respective components.[25] Incidentally, if $G_0 G_1 \cdots G_\lambda = \prod_p p^\infty$ and

$$(G_i, G_j) = 1 \text{ when } i \neq j$$

the usual local class field theory holds over k if the field of residue classes admits for each integer n exactly one cyclic extension of degree n, as in the preceding sections we have to restrict our considerations to fields and algebras whose degrees are relatively prime to the characteristic.

THE JOHNS HOPKINS UNIVERSITY.

---

[25] E. Steinitz, "Algebraische Theorie der Körper," *Crelle*, vol. 137 (1910).

THEORY OF MULTIGROUPS.*

By Melvin Dresher and Oystein Ore.

Various problems in non-commutative algebra lead naturally to the introduction of algebraic systems in which the operations are not one-valued. Let us only mention co-set expansions in groups with respect to non-normal subgroups or residue systems of one-sided ideals in rings. In the following we shall discuss the general theory of multigroups, i.e., algebraic systems satisfying all the axioms of a group except that the multiplication is multivalued.

The history of the theory of multigroups is very short. Multigroups (or hypergroups) were first defined by *Marty* (1934), who has studied their properties and applications in several communications.[1] Another contribution to this theory was made last year by H. S. *Wall*.[2] In addition one should mention two short notes by *Kuntzmann*[3] and some considerations by *Ore*[4] on co-set expansions in groups and their properties as multigroups.[4a]

In the following we have adopted the most general definition of a multigroup. We assume only the existence of solutions of linear relations and not the existence of two-sided units and two-sided inverses as in most previous investigations. We also find that the theory appears more elegant in this general form.

The first chapter contains a discussion of the definitions and representations of multigroups, together with examples. One also finds a theory of

---

* Received February 28, 1938.

[1] F. Marty, (1) "Sur une généralisation de la notion de groupe," *Huitième congrès des mathématiciens scandinaves*, Stockholm, 1934, pp. 45-49; (2) "Rôle de la notion d'hypergroupe dans l'étude des groupes non abéliens," *Comptes Rendus*, vol. 201 (Paris, 1935), pp. 636-638; (3) "Sur les groupes et hypergroupes attachés à une fraction rationelle," *Annales de l'école normale*, 3 sér., vol. 53 (1936), pp. 83-123.

[2] H. S. Wall, "Hypergroups," *American Journal of Mathematics*, vol. 59 (1937), pp. 77-98.

[3] J. Kuntzmann, (1) "Opérations multiformes. Hypergroupes," *Comptes Rendus*, vol. 204 (Paris, 1937), pp. 1787-1788; (2) "Homomorphie entre systèmes multiformes," *Comptes rendus*, vol. 205 (Paris, 1937), pp. 208-210.

[4] Oystein Ore, "Structures and group theory, I," *Duke Mathematical Journal*, vol. 3 (1937), pp. 149-174.

[4a] Added in proofs: Further contributions have appeared since the completion of this paper: L. W. Griffiths, "On hypergroups, multigroups and product systems," *American Journal of Mathematics*, vol. 60 (1938), pp. 345-354; M. Krasner, "Sur la primitivité des corps 𝔓-adiques," *Mathematica*, vol. 13 (1937), pp. 72-191.

scalars, i. e., elements for which the multiplication is unique. The concept of a scalar was first introduced by Wall. Here we extend his results in various ways.

In the second chapter we consider the properties of sub-multigroups. Usually they do not possess properties corresponding to those of subgroups in groups. The new concepts of *closure* and *reversibility* are introduced, and it is shown that for reversible sub-multigroups there exists co-set expansions and quotient multigroups. These investigations are connected with those of Marty and Kuntzmann on the homomorphisms of multigroups.

The last chapter deals with normal sub-multigroups and the decomposition theorems for multigroups. One can define two important types of normality in a multigroup. The weaker form for normality is sufficient to obtain analogues of the first and second theorem of isomorphism and the theorem of Jordan-Hölder. The stronger form of normality can be shown to be equivalent to the condition that the quotient multigroup shall be an ordinary group. Strongly normal sub-multigroups can only exist in multigroups which are homomorphic to ordinary groups. These investigations finally lead to a very interesting type of multigroups which we call *ultragroups.* These ultragroups are characterized by the property that they have composition series in which every quotient multigroup is an ordinary group.

CHAPTER I. General multigroups.

1. Definitions. A *multigroup* $\mathfrak{M}$ is an algebraic system with one operation called *multiplication.* This multiplication usually satisfies the ordinary group axioms except that *the product is not unique.* We enumerate the axioms as follows:

I. *The product.* If a and b are two elements of $\mathfrak{M}$ then the product $a \cdot b$ (simpler ab) is a sub-set of $\mathfrak{M}$

$$a \cdot b = (c_1, c_2, \cdot \cdot). \tag{1}$$

Let us observe explicitly that we make no assumptions on the number of elements in the product; it may be arbitrary and vary from product to product. We also extend the definition of a product to arbitrary subsets. If

$$A = (a_1, a_2, \cdot \cdot \cdot)$$
$$B = (b_1, b_2, \cdot \cdot \cdot)$$

then

$$A \cdot B = (\cdot \cdot \cdot, a_i b_j, \cdot \cdot \cdot) = \Sigma a_i b_j$$

is the set consisting of all elements of $\mathfrak{M}$ contained in some product $a_i b_j$. We shall write $A \supset a$, when a is an element of a set A.

II. *The associative law.* For any three elements of $\mathfrak{M}$ we have

$$(ab)c = a(bc) = abc.$$

These products have a meaning according to the definition of products of sub-sets.

III. *The quotient axiom.* To any two elements a and b there shall exist other elements x and y such that

$$ax \supset b, \qquad ya \supset b. \tag{2}$$

Any system whose elements satisfy these axioms we shall call a multigroup. If only the first relation (2) has a solution we shall call $\mathfrak{M}$ a *right multigroup* and if only the second has a solution we have a *left multigroup*.

An element e such that

$$ea \supset a$$

for all a in $\mathfrak{M}$ shall be called a *left unit*, and a *right unit* is defined analogously. A *unit* is an element e such that

$$ea \supset a, \qquad ae \supset a$$

for all a. If

$$e \cdot a = a$$

for all a, then e shall be called a *left scalar unit*. The *right scalar units* are defined in the same manner. If

$$ae = ea = a$$

for all a then e is a *scalar unit* or *absolute unit*.

THEOREM 1. *Let the multigroup $\mathfrak{M}$ contain a left (right) scalar unit e. If there exists any right (left) unit in $\mathfrak{M}$ then it is unique and equal to e and e is the only left (right) scalar unit of $\mathfrak{M}$. If $\mathfrak{M}$ contains an absolute unit then there are no other units.*

Proof. Let e_r be a right unit. Then

$$e \cdot e_r = e_r \supset e$$

and hence $e_r = e$. Any other left scalar unit would also have to be equal to e.

We shall say that a^{-1} is a *left inverse* of a when

$$a^{-1} \cdot a \supset e$$

where e is some unit element. The left inverse depends on the unit element e and even for a fixed e an element will usually have several inverses. We shall usually assume that by left inverses the corresponding e is a left unit. An inverse must always exist according to the quotient axiom. One defines *right inverses* in a similar manner and we shall associate a right inverse with a right unit element. A *two-sided inverse* a^{-1} has the properties

$$a^{-1} \cdot a \supset e, \quad a \cdot a^{-1} \supset e$$

where e is some (two-sided) unit element. Such inverses are not postulated in our multigroups.

One finds easily that if every product in a multigroup contains but one element the multigroup is a group.

A multigroup can always be imbedded in a *groupoid,* i. e. a system with a unique operation. Let us define a *complex* in $\mathfrak{M}$ to be a subset obtained by forming a product of a finite number of elements. The product of two such complexes is again a complex and the multiplication satisfies the associative law. One can also imbed a multigroup in a ring by writing (1) in the form

$$a \cdot b = c_1 + c_2 + \cdots.$$

By introducing rational integral coefficients to indicate the multiplicity of the occurring terms one can define the sum (and difference) of two complexes.

We shall also say that two multigroups $\mathfrak{M}$ and $\mathfrak{M}'$ are *isomorphic* if there exists a one-to-one correspondence between their elements such that if $a \rightleftarrows a'$ then $ab \supset c$ implies $a'b' \supset c'$.

The definition of a multigroup which we have given here and which we shall adopt in the following, corresponds to the definition of hypergroups by *Marty.* Almost all the investigations on multigroups by *Marty* and by *Wall* and *Kuntzmann* presuppose a stronger definition of a multigroup. Instead of the quotient axiom one assumes:

IV. There exists (two-sided) unit elements.

V. Every element has two-sided inverses.

Note: It might be mentioned that one can have multigroups in which every element is a unit. Such multigroups might be called *unit multigroups.*

Any multigroup satisfying these conditions may be called a *regular multigroup.* It is obvious that the conditions IV and V imply the existence of solutions of the relations (2). From

$$a^{-1} \cdot a \supset e, \quad a \cdot c^{-1} \supset e$$

where e is a unit follows namely

$$b \cdot a^{-1} \cdot c \supset b, \qquad a \cdot a^{-1} \cdot b \supset b$$

and hence $b \cdot a^{-1}$ contains an element y and $a^{-1} \cdot b$ an element x such that (2) holds.

2. Representations of multigroups. The multiplication in a finite multigroup may be represented by a Cayley square much in the same way as for a finite group. The only difference is that the various places in the table will contain several instead of a single element. We shall only illustrate this by an example of a multigroup of order 3.

	e	a	b
e	e	a, b	a, b
a	a	e, b	e, b
b	b	e, a	e, a

One finds that the associative law is satisfied. It is worth noting that $xa = xb$ for all x.

If $\mathfrak{M}$ is a multigroup of order n one can represent $\mathfrak{M}$ isomorphically by means of generalized permutations in which each index corresponds to a set of numbers. Let $a_1, \cdots, a_n$ be the elements of $\mathfrak{M}$. To an element b of $\mathfrak{M}$ we let correspond the symbols

$$P_b \rightarrow \begin{pmatrix} a_1, & \cdots, a_n \\ a_1 b, & \cdots, a_n b \end{pmatrix} \rightarrow \begin{pmatrix} 1, & 2, & \cdots, n \\ B_2, & B_2, & \cdots, B_n \end{pmatrix}$$

where B_i contains the indices of those elements of $\mathfrak{M}$ which are contained in $a_i b$. Such a substitution P_b is a unit if and only if B_i contains i for every i. If in the same manner

$$P_c \rightarrow \begin{pmatrix} a_1, & \cdots, a_n \\ a_1 c, & \cdots, a_n c \end{pmatrix} \rightarrow \begin{pmatrix} 1, & 2, & \cdots, n \\ C_1, & C_2, & \cdots, C_n \end{pmatrix}$$

then we define the product substitution

$$P_{bc} \rightarrow \begin{pmatrix} a_1, & \cdots, a_n \\ a_1 bc, & \cdots, a_n bc \end{pmatrix} \rightarrow \begin{pmatrix} 1, 2, \cdots, n \\ P_b : C_1, \ P_b : C_2, \cdots, P_b : C_n \end{pmatrix}.$$

By this definition one obtains an isomorphic representation of the multigroup.

One can also define a *matrix representation* of $\mathfrak{M}$ in a manner which corresponds to the ordinary regular representation of a group. One lets a correspond to a matrix

$$a \rightarrow A_a = (\delta_{i,j}^{(a)})$$

where

$$\delta_{i,j}^{(a)} = 1 \text{ when } aa_i \supset a_j$$
$$\delta_{i,j}^{(a)} = 0 \text{ otherwise.}$$

One finds that if

$$a \cdot b = (c_1, c_2, \cdots)$$

then one has for the corresponding matrices

$$A_a \cdot A_b = A_{c_1} + A_{c_2} + \cdots.$$

We shall not discuss further the properties of this representation.[5]

3. Scalars. We shall say that an element α is a *left scalar* when αx is a single element for all x. A *right scalar* is defined similarly. When αx and $x\alpha$ are single elements for all x we shall say that α is a *scalar*. Obviously the product of any two scalars of same type is again a scalar of the same type.

For the moment we shall consider only left scalars. We shall say that a product ab is a left scalar when abx is a single element for all x.

THEOREM 2. *When a product ab is a left scalar then a is a left scalar and $a \cdot b = x$ is a left scalar element.*

Proof. Since every element in $\mathfrak{M}$ is contained in some product bx we have that ay is single element for all y. Hence a is a left scalar and ab is a single element.

We shall say that the *cancellation law* holds for a left scalar α if any relation

$$\alpha x = \alpha y$$

implies $x = y$. In a finite multigroup the cancellation law is always satisfied because any element is representable in the form αx so that the two sets $\{x\}$ and $\{\alpha x\}$ contain the same number of elements.

We shall now prove:

THEOREM 3. *Let $\mathfrak{M}$ be a multigroup in which the cancellation law holds for left scalars. The existence of a left scalar then implies the existence of a left scalar unit.*

Proof. Let α be a left scalar and let us determine an e_a such that

$$\alpha \cdot e_a = \alpha. \tag{3}$$

Then one finds for an arbitrary x

---

[5] Compare Wall, *loc. cit.* (part I).

$$\alpha \cdot e_a \cdot x = \alpha x.$$

If now y is some element in $e_a x$ then $\alpha y = \alpha x$ or $y = x$. This gives

$$e_a \cdot x = x \tag{4}$$

for all x.

Under the assumptions of Theorem 3 one can also show:

THEOREM 4. *To each left scalar* α *there exists a unique left scalar inverse* α^{-1} *such that*

$$\alpha \cdot \alpha^{-1} = \alpha^{-1} \cdot \alpha = e_a. \tag{5}$$

Proof. The cancellation law gives the existence of a unique α^{-1} such that

$$\alpha \cdot \alpha^{-1} = e_a.$$

From this relation follows

$$\alpha \cdot \alpha^{-1}\alpha = e_a \cdot \alpha = \alpha = \alpha \cdot e_a$$

and hence we have

$$\alpha \cdot y = \alpha \cdot e_a$$

or $y = e_a$ for every y in $\alpha^{-1} \cdot \alpha$. Theorem 2 implies that α^{-1} is a left scalar.

When the cancellation law holds for left scalars one can strengthen Theorem 2 to:

THEOREM 5. *When* ab *is a left scalar then both* a *and* b *are left scalars.*

Proof. If bx contains c_1 and c_2 then $ac_1 = ac_2$ or $c_1 = c_2$. Hence bx is a single element.

THEOREM 6. *Let* $\mathfrak{M}$ *be a multigroup with a right unit* r, *in which the cancellation law holds for left scalars. If* $\mathfrak{M}$ *contains a left scalar then the set of all left scalars is a group with* r *as unit element.*

Proof. It follows from Theorem 1 that r is the only right unit and the only left scalar unit of $\mathfrak{M}$. The group property of the scalars follows from the two preceding theorems.

Let us now consider multigroups with (two-sided) scalars.[6]

THEOREM 6. *If a multigroup contains a scalar element then it contains an absolute unit element* e *and the set of all scalars is a group with* e *as unit element.*

---

[6] These are the only scalars considered by Wall.

Proof. Let α be the given scalar. We determine e_α such that

$$\alpha \cdot e_\alpha = \alpha.$$

Then one has for any x

$$x\alpha \cdot e_\alpha = x_\alpha.$$

Since every element of $\mathfrak{M}$ is representable in the form $x\alpha$ this means that e_α is a left scalar unit. In the same manner one establishes the existence of a right scalar unit and hence an absolute unit according to Theorem 1.

To prove that the scalars form a group let us determine α^{-1} such that

$$\alpha \cdot \alpha^{-1} = e.$$

Then

$$\alpha \cdot \alpha^{-1} \cdot \alpha = \alpha$$

and as before this implies

$$y \cdot \alpha^{-1}\alpha = y$$

for all y. Hence we have

$$\alpha^{-1} \cdot \alpha = \alpha \cdot \alpha^{-1} = e.$$

The inverse is a scalar according to Theorem 2. It is unique because any relation $\alpha x = \alpha y$ implies $x = y$ as one sees by multiplication with α^{-1}.

The group of scalars has been called the *nucleus* of $\mathfrak{M}$ by *Wall.*

THEOREM 7. *Let $\mathfrak{M}$ be a finite multigroup. If $\mathfrak{M}$ contains both a right and a left scalar then all scalars are two-sided.*

Proof. Since $\mathfrak{M}$ is finite the cancellation law holds and $\mathfrak{M}$ contains right and left scalar units according to Theorem 3. This implies the existence of an absolute unit element by Theorem 1. Furthermore we have for any left scalar α

$$\alpha^{-1} \cdot \alpha = \alpha \cdot \alpha^{-1} = e$$

and Theorem 2 shows that α is a scalar.

It is also easily seen:

THEOREM 8. *In any regular multigroup with an absolute unit element every left or right scalar is a scalar.*

4. Examples. The decompositions of ordinary groups furnish us with some of the most important examples of multigroups. Let us consider the quotient systems in groups as our first application.[7]

Let $\mathfrak{G}$ be a group and $\mathfrak{H}$ any subgroup. The co-sets $a\mathfrak{H}$ (or $\mathfrak{H}a$) form a set which we shall call the *quotient system* $\mathfrak{G}/\mathfrak{H}$ of $\mathfrak{G}$ with respect to $\mathfrak{H}$. This quotient system is a multigroup when one defines the product to be

---

[7] Compare O. Ore, *loc. cit.*, chap. I.

$$a\mathfrak{H} \cdot b\mathfrak{H} = (c_1\mathfrak{H}, c_2\mathfrak{H}, \cdots)$$

where the $c_i\mathfrak{H}$ are those co-sets which contain elements from the left-hand product. The multigroup is an ordinary group if and only if $\mathfrak{H}$ is normal in $\mathfrak{G}$. We shall assume that this is not the case. This is equivalent to saying that the normaliser group $\mathfrak{N}$ of $\mathfrak{H}$ in $\mathfrak{G}$ is a proper subgroup of $\mathfrak{G}$.

Under this assumption we find: *The only unit element in* $\mathfrak{G}/\mathfrak{H}$ *is* $\mathfrak{H}$ *and*

$$a\mathfrak{H} \cdot \mathfrak{H} = a\mathfrak{H}, \qquad \mathfrak{H} \cdot a\mathfrak{H} \supset a\mathfrak{H}.$$

If $b \cdot \mathfrak{H}$ is to be a r. h. inverse of $a\mathfrak{H}$ then one must have

$$a\mathfrak{H} \cdot b\mathfrak{H} \supset \mathfrak{H},$$

hence there must exist elements h_0 and h_1 in $\mathfrak{H}$ such that

$$ah_0b = h_1, \qquad b = h_1 \cdot a^{-1} \cdot h_0^{-1}.$$

It then follows easily: *Any inverse of* $a\mathfrak{H}$ *has the form* $ha^{-1} \cdot \mathfrak{H}$ *where* h *belongs to* $\mathfrak{H}$. *These inverses are all both r. h. and l. h. inverses, hence* $\mathfrak{G}/\mathfrak{H}$ *is regular multigroup.*

To determine the scalars in $\mathfrak{G}/\mathfrak{H}$ let us consider the condition for an equation

$$a\mathfrak{H} \cdot b\mathfrak{H} = c\mathfrak{H}. \tag{9}$$

to hold. One sees that c must have the form

$$c = ab \cdot h$$

and when this value is substituted in (9) one finds

$$b^{-1}\mathfrak{H}b = \mathfrak{H}.$$

Hence a relation (9) holds if and only if $b\mathfrak{H}$ belongs to the normaliser of $\mathfrak{H}$ in $\mathfrak{G}$. This shows: *The only right-hand scalars of* $\mathfrak{G}/\mathfrak{H}$ *are the co-sets of the normaliser of* $\mathfrak{H}$ *in* $\mathfrak{G}$. *There are no left-hand scalars.*

Another important type of multigroups is defined by the co-set expansion of a group $\mathfrak{G}$ with respect to two subgroups $\mathfrak{M}$ and $\mathfrak{N}$

$$\mathfrak{G} = \mathfrak{M}\,\mathfrak{N} + \mathfrak{M}a_2\mathfrak{N} + \cdots.$$

These double co-sets form a multigroup when one defines

$$\mathfrak{M}a\mathfrak{N} \cdot \mathfrak{M}b\mathfrak{N} = (\mathfrak{M}c_1\mathfrak{N}, \mathfrak{M}c_2\mathfrak{N}, \cdots)$$

where as before the right-hand side contains those co-sets which include elements from the l. h. product. One might denote this multigroup by $\mathfrak{M}\backslash\mathfrak{G}/\mathfrak{N}$. One can determine its units and inverses and one finds that it is a regular multigroup. The results are somewhat complicated and they shall not be reproduced here. It may be of interest to note that scalars only exist when the two groups are permutable.

In the simplest case where $\mathfrak{M}=\mathfrak{N}$ let us denote the multigroup $\mathfrak{N}\backslash\mathfrak{G}/\mathfrak{N}$ by $\mathfrak{G}/\!/\mathfrak{N}$. This multigroup has the absolute unit $\mathfrak{N}$. The inverse of $\mathfrak{N}a\mathfrak{N}$ is unique and equal to $\mathfrak{N}a^{-1}\mathfrak{N}$. Again the scalars are the co-sets belonging to the normaliser of $\mathfrak{N}$ in $\mathfrak{G}$.

CHAPTER 2. Co-set expansions.

1. Closed sub-multigroups. Let $\mathfrak{M}$ be a multigroup. We shall say that a sub-set $\mathfrak{A}$ is *multiplicatively closed* when it has the property that if a_1 and a_2 belong to $\mathfrak{A}$ then all elements contained in the product a_1a_2 also belong to $\mathfrak{A}$. We shall say that $\mathfrak{A}$ is a *sub-multigroup* of $\mathfrak{M}$, if it satisfies the axioms of a multigroup, i. e. if the relations

$$a_1x\supset a_2,\qquad ya_1\supset a_2$$

always have solutions in $\mathfrak{A}$.

When $\mathfrak{A}$ and $\mathfrak{B}$ are two sub-multigroups then the *cross-cut* $(\mathfrak{A},\mathfrak{B})$ of their common elements is multiplicatively closed. Usually $(\mathfrak{A},\mathfrak{B})$ is not a sub-multigroup, because if d_1 and d_2 are two of its elements the relations

$$d_1x\supset d_2,\qquad yd_1\supset d_2$$

need not have solutions in $(\mathfrak{A},\mathfrak{B})$.

This leads us to the consideration of more special types of sub-multigroups. We shall say that a sub-multigroup $\mathfrak{A}$ is *left closed* with respect to $\mathfrak{M}$ if for any two elements a_1 and a_2 in $\mathfrak{A}$ all the solutions y of the relation

$$ya_1\supset a_2 \tag{1}$$

lie in $\mathfrak{A}$. Similarly we say that $\mathfrak{A}$ is *right closed* when all the solutions of

$$a_1x\supset a_2 \tag{2}$$

lie in $\mathfrak{A}$. Finally $\mathfrak{A}$ is *closed* when all the solutions of (1) and (2) lie in $\mathfrak{A}$.

One sees immediately:

THEOREM 1. *The cross-cut* $(\mathfrak{A}, \mathfrak{B})$ *of two (left) closed sub-multigroups is void or a (left) closed sub-multigroup.*

The definition shows that a (left) closed sub-multigroup must contain all (left) units of $\mathfrak{M}$. Hence if $\mathfrak{M}$ contains left units the cross-cut of two left closed sub-multigroups is never void. In the following we shall usually assume the existence of such a unit element. Let us also observe that any (left) closed sub-multigroup contains all its (left) inverses.

The *union* $[\mathfrak{A}, \mathfrak{B}]$ of two sub-multigroups (or multiplicatively closed sub-sets) consists of all those elements which are contained in some product made up of factors from $\mathfrak{A}$ and $\mathfrak{B}$. The union is multiplicatively closed, but it is usually not a sub-multigroup even if $\mathfrak{A}$ and $\mathfrak{B}$ are closed. These remarks show that if the void set is included in the multiplicatively closed sets, one has:

THEOREM 2. *The set of all multiplicatively closed sub-systems form a structure.*

A similar theorem does usually not hold for the sub-multigroups.

One finds that in the multigroups defined by ordinary co-set expansions $\mathfrak{G}/\mathfrak{H}$ every sub-multigroup is closed. The same holds for the double co-set expansions $\mathfrak{G}/\!/\mathfrak{H}$.

2. Reversibility. Let $\mathfrak{M}$ be a multigroup. We shall say that a sub-multigroup $\mathfrak{A}$ of $\mathfrak{M}$ is *left reversible in* $\mathfrak{M}$ if any relation

$$a_1 m_1 \supset m_2$$

where a_1 belongs to $\mathfrak{A}$ implies the existence of an a_2 in $\mathfrak{A}$ such that

$$m_1 \subset a_2 m_2.$$

When m_2 is taken as an element in $\mathfrak{A}$ one finds that m_1 is an element of $\mathfrak{A}$, hence:

THEOREM 3. *Left reversibility implies right closure.*

Similarly we say that $\mathfrak{A}$ is *right reversible* in $\mathfrak{M}$, when any relation

$$m_1 a_1 \supset m_2$$

implies the existence of an a_2 in $\mathfrak{A}$ such that

$$m_1 \subset m_2 a_2.$$

Finally $\mathfrak{A}$ is *reversible* in $\mathfrak{M}$ when it is both left and right reversible. According to Theorem 3 a reversible sub-multigroup is closed.

We can now prove:

Theorem 4. *Let $\mathfrak{A}$ and $\mathfrak{B}$ be closed and $\mathfrak{A}$ (left) reversible in $\mathfrak{M}$. Then $(\mathfrak{A}, \mathfrak{B})$ is closed and (left) reversible in $\mathfrak{B}$.*

Proof. According to Theorem 1, $(\mathfrak{A}, \mathfrak{B})$ is closed. Furthermore since $\mathfrak{A}$ is left reversible any relation

$$d_1 b_1 \supset b_2$$

where d_1 belongs to $(\mathfrak{A}, \mathfrak{B})$, implies that

$$b_1 \subset a_2 \cdot b_2$$

and since $\mathfrak{B}$ is closed we have $a_2 = d_2$.

We are now able to prove:

Theorem 5. *The union of two reversible multigroups is a reversible multigroup.*

Proof. If c_1 and c_2 are arbitrary elements in $[\mathfrak{A}, \mathfrak{B}]$ then we shall first have to prove that any relations

$$(3) \qquad xc_1 \supset c_2, \qquad c_1 y \supset c_2$$

imply that x and y be in $[\mathfrak{A}, \mathfrak{B}]$. It is sufficient to consider the first of these relations. Each element of the union is contained in some product of factors from $\mathfrak{A}$ and $\mathfrak{B}$. Let us suppose that

$$(4) \qquad a_n b_n \cdot \cdot \cdot a_1 b_1 \supset c_1$$

is the product of shortest length containing c_1. When c_1 in (3) is contained in $\mathfrak{A}$ or $\mathfrak{B}$ our assertion is true by the reversibility of $\mathfrak{A}$ and $\mathfrak{B}$. Hence we may prove it in general by induction, assuming there exists solutions for all c_1 contained in products of shorter length. From (3) and (4) one obtains

$$x a_n b_n \cdot \cdot \cdot a_1 b_1 \supset c_2$$

and hence there exists an element d such that

$$x \cdot d \cdot b_1 \supset c_2, \qquad d \subset a_n b_n \cdot \cdot \cdot a_1.$$

This in turn implies the existence of an x_1 such that

$$x_1 \cdot b_1 \supset c_2, \qquad x_1 \subset xd.$$

The first of these relations show that x_1 belongs to $[\mathfrak{A}, \mathfrak{B}]$ and since d is contained in a product of shorter length than c_1 it follows from the second that x also belongs to $[\mathfrak{A}, \mathfrak{B}]$. The reversibility of $[\mathfrak{A}, \mathfrak{B}]$ follows in a similar manner.

Now let us define further: Let $\mathfrak{M}$ be a multigroup with (left, right) units. We shall say that $\mathfrak{M}$ is (left, right) *reversible in itself* when any relation

$$m_1 \cdot m_2 \supset m_3$$

implies the existence of a left inverse m_1^{-1} and a right inverse m_2^{-1} such that

$$m_2 \subset m_1^{-1} \cdot m_3, \qquad m_1 \subset m_3 \cdot m_2^{-1}.$$

From this definition follows:

THEOREM 6. *Let $\mathfrak{M}$ be (left) reversible in itself. Then every closed sub-multigroup is left reversible in $\mathfrak{M}$.*

When this theorem is combined with Theorems 4 and 5 one obtains:

THEOREM 7. *Let $\mathfrak{M}$ be a multigroup which is reversible in itself. Then the set of all closed sub-multigroups of $\mathfrak{M}$ form a structure.*

Let us observe that in ordinary groups the co-set expansions $\mathfrak{G}/\mathfrak{H}$ and $\mathfrak{G}/\!/\mathfrak{H}$ define multigroups which are reversible in themselves.

3. Co-set expansions. Let $\mathfrak{A}$ denote some sub-multigroup or even only a multiplicatively closed sub-set of $\mathfrak{M}$. By $\mathfrak{A}m$ we shall denote the set of all elements contained in some product $a \cdot m$ where a belongs to $\mathfrak{A}$. We shall call $\mathfrak{A}m$ a *left co-set* of $\mathfrak{M}$ with respect to $\mathfrak{A}$.

From now on we shall suppose that $\mathfrak{A}$ is left reversible in $\mathfrak{M}$. Since any relation

$$a_1 m_1 \supset m_2$$

implies

$$m_1 \subset a_2 \cdot m_2$$

or

$$\mathfrak{A}m_1 = \mathfrak{A}m_2$$

we can state:

THEOREM 8. *Any element in a co-set $\mathfrak{A}m$ generates the same co-set and a co-set contains its generating element.*

This gives us the main theorem:

THEOREM 9. *Let $\mathfrak{A}$ be a left reversible multiplicatively closed sub-system of a multigroup $\mathfrak{M}$. Then $\mathfrak{M}$ has a unique co-set expansion*

$$\mathfrak{M} = \mathfrak{A}m_1 + \mathfrak{A}m_2 + \cdots \tag{5}$$

where the co-sets $\mathfrak{A}m$ *are disjoint sets.*

Proof. Every element in $\mathfrak{M}$ belongs to some co-set and two co-sets with a common element are identical according to Theorem 3.

Two co-sets will usually not contain the same number of elements. Let us also observe that the left reversibility of $\mathfrak{A}$ implies that $\mathfrak{A}$ is a right multigroup which is right closed.

The existence of double co-set expansions may be derived under similar conditions. We prove:

THEOREM 10. *Let* $\mathfrak{A}$ *be a left reversible and* $\mathfrak{B}$ *a right reversible multiplicatively closed sub-set of* $\mathfrak{M}$. *Then* $\mathfrak{M}$ *has a double co-set expansion*

$$\mathfrak{M} = \mathfrak{A}m_1\mathfrak{B} + \mathfrak{A}m_2\mathfrak{B} + \cdots \tag{6}$$

where the co-sets $\mathfrak{A}m\mathfrak{B}$ *are disjoint sets.*

Proof. As before it suffices to prove that any element in a double co-set generates the same co-set. Now if

$$amb \supset m'$$

then there exists an $\bar{m}$ such that

$$\bar{m} \subset a \cdot m, \qquad m \subset a_1 \cdot \bar{m}$$

and such that

$$\bar{m}b \supset m', \qquad \bar{m} \subset m'b_1.$$

This gives

$$a_1 m' b_1 \supset a_1 \bar{m} \supset m$$

and our assertion is proved.

As a special case we have:

THEOREM 11. *Let* $\mathfrak{A}$ *be a sub-multigroup reversible in* $\mathfrak{M}$. *Then there exists a unique double co-set expansion*

$$\mathfrak{M} = \mathfrak{A}m_1\mathfrak{A} + \mathfrak{A}m_2\mathfrak{A} + \cdots. \tag{7}$$

4. Quotient multigroups.[8] Let us assume that $\mathfrak{A}$ is a left reversible sub-multigroup of $\mathfrak{M}$. Then $\mathfrak{A}$ is seen to be one of the co-sets (5). These co-sets $\mathfrak{A}m$ may be considered to be the elements of a new multigroup which we shall call the *quotient multigroup* of $\mathfrak{M}$ with respect to $\mathfrak{A}$ and denote by $\mathfrak{M}/\mathfrak{A}$. We define the product of two co-sets to be

---

[8] Compare Marty (3).

$$\mathfrak{A}m_1 \cdot \mathfrak{A}m_2 = (\mathfrak{A}n_1, \mathfrak{A}n_2, \cdots)$$

where the co-sets on the right are those which contain elements from the left-hand complexes. To prove that $\mathfrak{M}/\mathfrak{A}$ is a multigroup one must show that the relations

$$\mathfrak{A}m_1 \cdot \mathfrak{A}x \supset \mathfrak{A}m_2, \qquad \mathfrak{A}y \cdot \mathfrak{A}m_1 \supset \mathfrak{A}m_2$$

always have solutions. One sees however that it is sufficient to choose x and y such that

$$m_1x \supset m_2, \qquad ym_1 \supset m_2.$$

THEOREM 11. *Let $\mathfrak{A}$ be a left reversible sub-multigroup of $\mathfrak{M}$. The left co-set expansion of $\mathfrak{M}$ with respect to $\mathfrak{A}$ defines a quotient multigroup $\mathfrak{M}/\mathfrak{A}$ with the left scalar unit $\mathfrak{A}$.*

One might add that when $\mathfrak{M}$ contains a right unit then it must be contained in $\mathfrak{A}$ since $\mathfrak{A}$ is right closed, and hence $\mathfrak{A}$ is a right unit. It is the only right unit according to Theorem 1, chap. 1. When $\mathfrak{A}$ is closed, then $\mathfrak{A}$ is the only right or left unit.

Now let us consider any multigroup $\mathfrak{B}$ between $\mathfrak{M}$ and $\mathfrak{A}$, i. e.

$$\mathfrak{M} \supset \mathfrak{B} \supset \mathfrak{A}.$$

One can expand $\mathfrak{B}$ in co-sets with respect to $\mathfrak{A}$ and these co-sets obviously form a sub-multigroup $\mathfrak{B}/\mathfrak{A}$ of $\mathfrak{M}/\mathfrak{A}$. One finds without difficulty that if $\mathfrak{B}$ is closed or reversible in $\mathfrak{M}$ in some sense then $\mathfrak{B}/\mathfrak{A}$ has the same property in $\mathfrak{M}/\mathfrak{A}$.

Now let us consider the converse correspondence. Let $\mathfrak{B}^*$ be a sub-multigroup of $\mathfrak{M}/\mathfrak{A}$ containing $\mathfrak{A}$. We can then show that $\mathfrak{B}^*$ when considered as a set of elements in $\mathfrak{M}$ will be a multigroup $\mathfrak{B}$ containing $\mathfrak{A}$. To verify this we shall have to show that the relations

$$xb_1 \supset b_2, \qquad b_1y \supset b_2 \tag{8}$$

have solutions in $\mathfrak{B}$ when b_1 and b_2 are elements of this set.

To find an x satisfying the first relation (8) let us observe that since $\mathfrak{B}^*$ is a multigroup there exists an x_1 in $\mathfrak{B}$ such that

$$\mathfrak{A}x_1 \cdot \mathfrak{A}b_1 \supset \mathfrak{A}b_2.$$

This implies

$$a_1x_1a_2 \cdot b_1 \supset b_2$$

and hence $a_1x_1a_2$ contains an element x such that the first relation (8) holds.

To find a solution y of the second relation (8) let us determine z_1 in $\mathfrak{B}$ such that

$$\mathfrak{A}b_1 \cdot \mathfrak{A}z_1 \supset \mathfrak{A}.$$

From this relation follows that one can determine a z_a in $\mathfrak{B}$ such that

$$b_1 \cdot z_a \supset a$$

for an arbitrary a in $\mathfrak{A}$. Since $\mathfrak{A}b_2$ contains b_2 it follows by right multiplication with b_2 that a solution y of (8) in $\mathfrak{B}$ exists.

Let us observe that one can also consider the co-sets of a double co-set expansion (6) or (7) as the elements of a new multigroup $\mathfrak{A}\backslash\mathfrak{M}/\mathfrak{B}$ or $\mathfrak{M}/\!/\mathfrak{A}$. Similar results can be derived for such multigroups.

5. Homomorphisms. We shall say that a multigroup $\mathfrak{M}$ is *homomorphic* to another multigroup $\mathfrak{M}^*$ when there exists a correspondence $m \rightarrow m^*$ between the elements of the two systems such that when[9]

$$a \cdot b \supset c$$

then

$$a^* \cdot b^* \supset c^*.$$

Furthermore we suppose that every element of $\mathfrak{M}^*$ is the image of some element of $\mathfrak{M}$.

Obviously the co-set correspondence

$$m \rightarrow \mathfrak{A}m = m^*$$

which we have studied, defines a homomorphism of $\mathfrak{M}$ upon $\mathfrak{M}/\mathfrak{A}$. Also the double co-set expansions determine multigroups homomorphic to $\mathfrak{M}$.

$$m \rightarrow \mathfrak{A}m\mathfrak{B}, \qquad n \rightarrow \mathfrak{A}n\mathfrak{A}.$$

We shall study a few of the properties of such homomorphisms.

THEOREM 12. *All those elements of $\mathfrak{M}$ which correspond to a left scalar unit of $\mathfrak{M}^*$ form a right multigroup which is right closed. All those elements of $\mathfrak{M}$ which correspond to an absolute unit element of $\mathfrak{M}^*$ form a closed sub-multigroup.*

Proof. First if e^* is an idempotent in $\mathfrak{M}^*$ i. e. an element such that

$$e^{*2} = e^*,$$

---

[9] Compare Marty (3) and Kuntzmann (2).

then the elements in $\mathfrak{M}$ corresponding to e^* must form a set $\mathfrak{A}$ which is multiplicatively closed. Now let e^* be a left scalar unit, and a_1 and a_2 two elements of $\mathfrak{A}$. From a relation $a_1 x \supset a_2$ follows immediately $x^* = e^*$ and x also belongs to $\mathfrak{A}$.

In order to derive further properties of the homomorphism it is necessary to make assumptions on the inverse correspondence from $\mathfrak{M}^*$ to $\mathfrak{M}$. Let us say that there exists a *strong* (*left*) *homomorphism* between $\mathfrak{M}$ and $\mathfrak{M}^*$ when the following condition is satisfied.

Let

$$a^* b^* \supset c^*.$$

Then to any b_0 and any c_0 corresponding to b^* and c^* respectively there exists an a corresponding to a^* such that

$$a \cdot b_0 \supset c_0.$$

One sees easily that the correspondence between $\mathfrak{M}$ and a left co-set expansion $\mathfrak{M}/\mathfrak{A}$ is a strong left homomorphism. This need not be true for the two-sided co-set expansion $\mathfrak{M}/\!/\mathfrak{A}$.

We can now prove:

THEOREM 13. *Let $\mathfrak{M}$ be strongly left homomorphic to $\mathfrak{M}^*$ and let $\mathfrak{A}$ be the right closed, right sub-multigroup which consists of the elements corresponding to a left scalar unit e^* of $\mathfrak{M}^*$. Then $\mathfrak{A}$ is left reversible in $\mathfrak{M}$ and $\mathfrak{M}^*$ is isomorphic to the quotient multigroup $\mathfrak{M}/\mathfrak{A}$ defined by the left co-set expansion of $\mathfrak{M}$ with respect to $\mathfrak{A}$.*

Proof. Let us suppose that two elements m_1 and m_2 have the same image. From

$$e^* \cdot m^*_1 = m^*_1 = m^*_2 = e^* \cdot m^*_2$$

follows on account of the strong homomorphism

$$a_1 \cdot m_1 \supset m_2, \qquad m_1 \subset a_2 \cdot m_2$$

where a_1 and a_2 belong to $\mathfrak{A}$. Hence all m with the same image belong to the same left co-set $\mathfrak{A}m$ and two elements of the same co-set obviously have the same image.

If one wishes to characterize also the double co-set expansions one must make a slightly different condition on the homomorphism: If

$$a^* \cdot b^* \cdot c^* \supset d^*$$

then it must be possible to determine to any b_0 and d_0 such an a and such a c that

$$a \cdot b_0 \cdot c \supset d_0.$$

When this condition is satisfied and $\mathfrak{A}$ is the closed sub-multigroup of $\mathfrak{M}$ which corresponds to an absolute unit element of $\mathfrak{M}^*$, then two elements of $\mathfrak{M}$ with the same image belong to the same double co-set with respect to $\mathfrak{A}$.

6. Some special multigroups. Let us conclude this chapter by a short discussion of the properties of those multigroups which are obtained from co-set expansions $\mathfrak{G}/\mathfrak{H}$ in groups. The multigroup defined by the left co-set expansion of a group with respect to a subgroup has a single unit element and is reversible. In addition it satisfies the following condition:

Cancellation law. Let $\mathfrak{M}$ be a multigroup with a single unit element e. We say that $\mathfrak{M}$ satisfies the *(left) cancellation law* when any relation

$$ab \supset b$$

implies $a = e$.

We can prove:

THEOREM 8. *A multigroup satisfying the (left) cancellation law is regular*

Proof. We shall have to prove that every left inverse is a right inverse and conversely. From

$$m^{-1} \cdot m \supset e$$

follows

$$m \cdot m^{-1} \cdot m \supseteq m$$

or

$$m \cdot m^{-1} \supset e.$$

From this last relation follows conversely

$$m^{-1} \cdot m \cdot m^{-1} \supseteq m^{-1}$$

or

$$m^{-1} \cdot m \supset e.$$

We can also prove:

THEOREM 9. *A multigroup satisfying the left cancellation law is left reversible.*

Proof. Let us suppose that

$$ab \supset c. \tag{9}$$

We determine x such that $xc \supset b$ and find

$$x \cdot ab \supset xc \supset b.$$

Hence one concludes $x = a^{-1}$ and (9) implies

$$a^{-1} \cdot c \supset b.$$

The double co-set expansions of a group $\mathfrak{G}//\mathfrak{H}$ have still simpler properties. They belong to the following type of multigroups:

Completely regular multigroups. A multigroup is said to be *completely regular* when it contains a single unit element e and every element has a unique inverse such that

$$m \cdot m^{-1} \supset e, \qquad m^{-1} \cdot m \supset e.$$

For such multigroups one can prove:

Theorem 10. *In a completely regular, reversible multigroup the right and left co-set expansions contain the same number of co-sets.*

Proof. Let (5) denote the left co-set expansion of $\mathfrak{M}$ with respect to a sub-multigroup $\mathfrak{A}$. We want to show that

$$\mathfrak{M} = m_1^{-1} \cdot \mathfrak{A} + m_2^{-1} \cdot \mathfrak{A} + \cdots \tag{10}$$

is a right co-set expansion. We show first that every element x belongs to some co-set (10). Since x^{-1} belongs to some left co-set, we have

$$am_i \supset x^{-1}$$

which successively implies

$$am_ix \supset e, \qquad m_ix \supset a^{-1}, \qquad x \subset m_i^{-1} \cdot a^{-1}.$$

Secondly no two co-sets (10) are identical since

$$m_i^{-1} \cdot a \supset m_j^{-1}$$

is seen to imply

$$m_j \subset a^{-1} \cdot m_i$$

against assumption.

Let us mention finally that under some very restrictive conditions one can also prove a theorem analogous to the theorem of Lagrange, that the order of a subgroup divides the order of the group.

CHAPTER 3. Decomposition theorems.

1. Normal sub-multigroups and quotient systems. We shall now introduce the concept of a *normal sub-multigroup* or more general, a *normal set* in a multigroup $\mathfrak{M}$. A set $\mathfrak{A}$ is *normal* in $\mathfrak{M}$, if for any m in $\mathfrak{M}$ we have

$$m \cdot \mathfrak{A} = \mathfrak{A} \cdot m$$

i. e. every element contained in a product $m \cdot a_1$ is also contained in some product $a_2 \cdot m$.

For a normal sub-multigroup left or right closure implies closure. Similarly left or right reversibility implies reversibility. Since we usually wish to be able to perform a co-set decomposition of $\mathfrak{M}$ with respect to a normal sub-multigroup $\mathfrak{A}$, we shall in most cases suppose also that $\mathfrak{A}$ is reversible in $\mathfrak{M}$.

One finds easily:

THEOREM 1. *Let $\mathfrak{A}$ be a normal reversible sub-multigroup of $\mathfrak{M}$. The quotient multigroup $\mathfrak{M}/\mathfrak{A}$ is strongly homomorphic to $\mathfrak{M}$ and has the absolute unit $\mathfrak{A}$. Conversely by any strong homomorphism $\mathfrak{M} \rightarrow \mathfrak{M}^*$ those elements of $\mathfrak{M}$ which correspond to the absolute unit element of $\mathfrak{M}^*$ form a normal reversible sub-multigroup $\mathfrak{A}$ of $\mathfrak{M}$ such that $\mathfrak{M}/\mathfrak{A}$ is isomorphic to $\mathfrak{M}^*$.*

Proof. According to the results of § 5 chap. 2 it is only necessary to prove that if e^* is an absolute unit element of $\mathfrak{M}^*$ then those elements $\mathfrak{A}$ of $\mathfrak{M}$ which correspond to e^* form a normal sub-multigroup. Since for any element of $\mathfrak{M}^*$

$$e^* \cdot m^* = m^* = m^* \cdot e^*$$

it follows that all elements with the same image must belong to the co-set $\mathfrak{A}m$ and the co-set $m\mathfrak{A}$.

When $\mathfrak{A}$ is normal and reversible in $\mathfrak{M}$ then the right and left co-set expansions are identical and the quotient system $\mathfrak{M}/\mathfrak{A}$ is uniquely determined. We prove next:

THEOREM 2. *Let $\mathfrak{A}$ be a normal reversible sub-multigroup of $\mathfrak{M}$. Then there exists a one-to-one correspondence between the normal reversible sub-multigroups of $\mathfrak{M}/\mathfrak{A}$ and the normal reversible sub-multigroups of $\mathfrak{M}$ containing $\mathfrak{A}$.*

Proof. It is seen immediately from the definition of $\mathfrak{M}/\mathfrak{A}$ that every normal reversible sub-multigroup $\mathfrak{B}$ between $\mathfrak{M}$ and $\mathfrak{A}$ corresponds to a unique normal reversible sub-multigroup $\mathfrak{B}^*$ of $\mathfrak{M}/\mathfrak{A}$. To prove the converse let $\mathfrak{B}^*$

be a reversible normal sub-multigroup of $\mathfrak{M}/\mathfrak{A}$. Since $\mathfrak{B}^*$ is closed, it contains the unit element $\mathfrak{A}$, hence it corresponds to some multigroup between $\mathfrak{A}$ and $\mathfrak{M}$. From

$$\mathfrak{B} \cdot \mathfrak{A}m = \mathfrak{A}m \cdot \mathfrak{B}$$

we also find

$$\mathfrak{B} \cdot m = m \cdot \mathfrak{B}$$

for any m, hence $\mathfrak{B}$ is normal in $\mathfrak{M}$. To prove finally that any relations

$$b_1 \cdot m_1 \supset m_2, \qquad m_1 \cdot b_2 \supset m_2$$

imply

$$m_1 \subset b_1 \cdot m_2, \qquad m_1 \subset m_2 \cdot b_2$$

it is only necessary to consider the corresponding relations which hold for the co-sets containing these elements.

2. The theorem of isomorphism. We shall now deduce various structural relations for normal sub-multigroups.[10]

THEOREM 3. *Let $\mathfrak{M}$ be a multigroup with units of some kind, and let $\mathfrak{A}$ and $\mathfrak{B}$ be normal reversible sub-multigroups. Then the union $[\mathfrak{A}, \mathfrak{B}]$ is also a normal reversible sub-multigroup and all its elements are contained in some product $a \cdot b$.*

Proof. It follows from the definition of normality that it is only necessary to assume that $\mathfrak{A}$ is normal in order that every element of the union be of the form a or b or contained in a product $a \cdot b$. When $\mathfrak{A}$ contains a left or $\mathfrak{B}$ a right unit of $\mathfrak{M}$, then every element of the union is contained in a product $a \cdot b$. This is always the case when $\mathfrak{M}$ contains some unit element and $\mathfrak{A}$ and $\mathfrak{B}$ are reversible. The normality of the union follows from

$$m \cdot \mathfrak{A} \cdot \mathfrak{B} = \mathfrak{A} \cdot m \cdot \mathfrak{B} = \mathfrak{A} \cdot \mathfrak{B} \cdot m$$

and the reversibility has been proved in Theorem 5, chap. 2.

THEOREM 4. *Let $\mathfrak{M}$ contain some unit element and let $\mathfrak{A}$ and $\mathfrak{B}$ be normal reversible sub-multigroups. Then the cross-cut $(\mathfrak{A}, \mathfrak{B})$ is normal and reversible in $\mathfrak{A}$ and $\mathfrak{B}$.*

Proof. Since $\mathfrak{M}$ contains unit elements the cross-cut $\mathfrak{D} = (\mathfrak{A}, \mathfrak{B})$ is not void. Furthermore $\mathfrak{D}$ is reversible in $\mathfrak{A}$ and in $\mathfrak{B}$ according to Theorem 4, chap. 2. To prove the normality of $\mathfrak{D}$ in $\mathfrak{B}$ let us observe that since $\mathfrak{D}$ is contained in $\mathfrak{A}$ we have

---

[10] See also Marty (3).

$$b\mathfrak{D} \subset \mathfrak{A}b$$

and since $\mathfrak{B}$ is closed one finds

$$b\mathfrak{D} \subset \mathfrak{D}b$$

and the equality is obtained by the converse argument.

THEOREM 5. *The Dedekind relation. Let* $\mathfrak{M}$ *contain some unit element and let* $\mathfrak{A}$, $\mathfrak{B}$, $\mathfrak{C}$ *be normal reversible sub-multigroups such that* $\mathfrak{C} \supset \mathfrak{A}$. *Then*

$$(1) \qquad (\mathfrak{C}, [\mathfrak{A}, \mathfrak{B}]) = [\mathfrak{A}, (\mathfrak{C}, \mathfrak{B})].$$

Proof. The right-hand side of (1) is obviously always contained in the left-hand side. To prove that the left-hand side is contained in the right-hand side we need actually only that $\mathfrak{A}$ is normal and reversible. Any element c contained in the left-hand side then satisfies a relation

$$c \subset a \cdot b$$

and since $\mathfrak{A}$ is reversible it follows that $b = d$ is contained in $\mathfrak{C}$. Hence we have

$$c \subset a \cdot d$$

where d is contained in $(\mathfrak{B}, \mathfrak{C})$.

Finally we prove the *theorem of isomorphism:*

THEOREM 6. *Let* $\mathfrak{A}$ *be normal and reversible and* $\mathfrak{B}$ *closed in the union* $[\mathfrak{A}, \mathfrak{B}]$. *We suppose further that the cross-cut* $(\mathfrak{A}, \mathfrak{B})$ *is not void. Then* $(\mathfrak{A}, \mathfrak{B})$ *is normal and reversible in* $\mathfrak{B}$ *and there exists an isomorphism between the quotient systems*

$$(2) \qquad [\mathfrak{A}, \mathfrak{B}]/\mathfrak{A} \cong \mathfrak{B}/(\mathfrak{A}, \mathfrak{B}).$$

Proof. It follows from Theorem 4, chap. 2 that $(\mathfrak{A}, \mathfrak{B}) = \mathfrak{D}$ is reversible in $\mathfrak{B}$. The normality of $\mathfrak{D}$ in $\mathfrak{B}$ follows as in Theorem 4. Now let

$$(3) \qquad \mathfrak{B} = \mathfrak{D} + b_2\mathfrak{D} + \cdots$$

be the co-set expansion of $\mathfrak{B}$ with respect to $\mathfrak{D}$. We can then show that

$$(4) \qquad [\mathfrak{A}, \mathfrak{B}] = \mathfrak{A} + b_2\mathfrak{A} + \cdots$$

is the co-set expansion of the union with respect to $\mathfrak{A}$. First every element of $[\mathfrak{A}, \mathfrak{B}]$ must belong to one of the co-sets (4) since every b belongs to some co-set (3). Secondly any relation

$$b_i\mathfrak{A} = b_j\mathfrak{A} \qquad i \neq j$$

would imply that

$$b_i \subset b_j \cdot a$$

where a belongs both to $\mathfrak{A}$ and $\mathfrak{B}$, hence to $\mathfrak{D}$ since $\mathfrak{B}$ is closed. This is impossible however since b_i and b_j belong to different co-sets (3). Finally one sees that the correspondence establishes the isomorphism (2).

3. The theorem of Jordan-Hölder. The preceding results are sufficient to prove the analogue of the theorem of *Jordan-Hölder* for multigroups. A chain of sub-multigroups

$$\mathfrak{A} \supset \mathfrak{A} \supset \cdots \supset \mathfrak{A}_n = \mathfrak{B} \tag{5}$$

shall be called a *composition series* when each $\mathfrak{A}_i$ is normal and reversible in the preceding. We shall call (5) a *maximal composition series,* when no further terms can be intercalated such that the series remains a composition series.

THEOREM 7. *Any two maximal composition series between $\mathfrak{A}$ and $\mathfrak{B}$ have the same length and the quotient multigroups defined by consecutive terms in one series are isomorphic to those defined by the other in some order.*

Proof. The proof follows along the lines of the ordinary proof of the theorem of Jordan-Hölder using induction with regard to the length of the chain. If (5) and

$$\mathfrak{A} \supset \mathfrak{B}_1 \supset \cdots \supset \mathfrak{B}_m = \mathfrak{B}$$

are the two chains, then one compares them to the two chains

$$\mathfrak{A} \supset \mathfrak{A}_1 \supset (\mathfrak{A}_1, \mathfrak{B}_1) \supset \cdots \supset \mathfrak{B}$$
$$\mathfrak{A} \supset \mathfrak{B}_1 \supset (\mathfrak{A}_1, \mathfrak{B}_1) \supset \cdots \supset \mathfrak{B}$$

for which the theorem is true according to Theorem 6 and the induction assumption.

One can also prove the theorem of Jordan-Hölder in the stronger form of a refinement theorem corresponding to the theorem of Schreier-Zassenhaus for groups. This requires the proof of a lemma analogous to the lemma of Zassenhaus. We shall omit these considerations here since they do not add materially to the theory of multigroups.

4. Strong normality. In the preceding we have used a definition of normality which is as weak as possible, but still sufficient to prove the theorem of Jordan-Hölder. There are however other stronger types of normality with interesting properties.

One slight strengthening of our definition of normality is the following: The sub-multigroup $\mathfrak{A}$ shall be said to be normal in $\mathfrak{M}$ when any product $a_1 \cdot m$ can be written in the form $m \cdot a_2$ and conversely.

One can prove a few more properties of normal sub-multigroups on the basis of this definition. A more important type of normality may be defined as follows:

Strong normality. A sub-multigroup $\mathfrak{A}$ of $\mathfrak{M}$ is *strongly normal* when it contains left and right unit elements of $\mathfrak{M}$ and satisfies the following condition: To each m_1 there exists some left inverse m_1^{-1} such that

$$m_1\mathfrak{A}m_1^{-1} \subset \mathfrak{A} \tag{6}$$

and to each m_2 some right inverse m_2^{-1} such that

$$m_2^{-1} \cdot \mathfrak{A} \cdot m_2 \subset \mathfrak{A} \tag{7}$$

Through multiplication of (6) and (7) by m_1 and m_2 one finds

$$m\mathfrak{A} = \mathfrak{A}m \tag{8}$$

for all m, so that strong normality implies normality in the sense we have used in the preceding. We can now prove:

THEOREM 8. *Strong normality implies reversibility.*

Proof. We shall show first that a strongly normal sub-multigroup $\mathfrak{A}$ is closed. This requires that any x and y such that

$$x \cdot a_1 \supset a_2, \qquad a_1 \cdot y \supset a_2 \tag{9}$$

belong to $\mathfrak{A}$. On account of the normality (8) it is sufficient to show that the second relation (9) implies that y belongs to $\mathfrak{A}$. According to (6) there exists a left inverse y^{-1} of y such that

$$y \cdot y^{-1} \subset \mathfrak{A}. \tag{10}$$

From the second relation in (9) follows

$$\mathfrak{A} \supset a_1 \cdot y \cdot y^{-1} \supset a_2 \cdot y^{-1}.$$

When this relation is multiplied by a left inverse of a_2 it follows that y^{-1} belongs to $\mathfrak{A}$. From (10) one concludes in the same manner that y belongs to $\mathfrak{A}$.

To prove the reversibility of $\mathfrak{A}$ we shall have to show that a relation

$$a_1 \cdot m_1 \supset m_2 \tag{11}$$

implies the existence of an a_2 in $\mathfrak{A}$ such that

$$m_1 \subset a_2 \cdot m_2.$$

In any case we can find an element z such that

$$z \cdot m_2 \supset m_1$$

and this implies according to (11) that

$$a_1 z \cdot m_2 \supset m_2.$$

When this relation is multiplied by a suitable left inverse of m_2 we find as in (10)

$$a_1 \cdot z \cdot \mathfrak{A} \supset \mathfrak{A}$$

and from the closure of $\mathfrak{A}$ one easily concludes that z belongs to $\mathfrak{A}$.

We can now prove further:

THEOREM 9. *All right and left inverses of an element m belong to the same co-set with respect to a strongly normal sub-multigroup, $\mathfrak{A}$.*

Proof. Let $\bar{m}$ be any left inverse of m and m^{-1} the left inverse for which

$$m \cdot m^{-1} \subset \mathfrak{A}.$$

Then

$$\bar{m} \cdot m \cdot m^{-1} \subset \bar{m}\mathfrak{A}$$

or

$$m^{-1} \subset \bar{m}\mathfrak{A}.$$

This shows that all left inverses of m belong to the same co-set $m^{-1} \cdot \mathfrak{A}$. But the right inverses of m must also belong to this co-set since

$$m \cdot m^{-1} \supset a_1$$

hence a suitable inverse a_2 of a gives

$$m \cdot m^{-1} \cdot a_2 \supset e$$

so that $m^{-1}a_2$ contains a right inverse of m.

Theorem 9 implies:

Theorem 10. *The conditions of strong normality are equivalent to the relation*

$$m\mathfrak{A}m^{-1} = \mathfrak{A} \tag{12}$$

to hold for all m *and all inverses* m^{-1}.

It also shows that

$$m_1^{-1} \cdot m_1 \subset \mathfrak{A}, \qquad m_2 \cdot m_2^{-1} \subset \mathfrak{A}$$

for all inverses m_1^{-1} and m_2^{-1}. Furthermore if

$$m_1 x \supset m_2, \qquad y m_1 \supset m_2$$

then one finds that any other solutions of these relations belong to the co-sets $x\mathfrak{A}$ and $y\mathfrak{A}$ respectively.

5. Quotient groups. The preceding investigations on strong normality enables us to prove certain characteristic properties of strongly normal submultigroups. The following theorem is of great interest:

Theorem 11. *When* $\mathfrak{A}$ *is strongly normal in* $\mathfrak{M}$ *then the quotient multigroup* $\mathfrak{M}/\mathfrak{A}$ *is an ordinary group.*

Proof. According to Theorem 1 the quotient multigroup $\mathfrak{M}/\mathfrak{A}$ has the absolute unit element $\mathfrak{A}$. Theorem 9 shows furthermore that

$$m \cdot \mathfrak{A} \cdot m^{-1} \cdot \mathfrak{A} = \mathfrak{A}.$$

This relation shows according to Theorem 2, chap. 1 that $m\mathfrak{A}$ is a left scalar in $\mathfrak{M}/\mathfrak{A}$. In the same manner follows that $m\mathfrak{A}$ is a right scalar hence every element of $\mathfrak{M}/\mathfrak{A}$ is a scalar and the quotient multigroup is a group.

The following theorem supplements Theorem 11:

Theorem 12. *The necessary and sufficient condition that a multigroup* $\mathfrak{M}$ *be homomorphic to a group* G *is that* $\mathfrak{M}$ *contain a strongly normal submultigroup* $\mathfrak{A}$ *such that the group* $\mathfrak{M}/\mathfrak{A}$ *is isomorphic to* G.

Proof. Let us observe first that the theorem only assumes homomorphism in the weakest sense, but that the proof shows that $\mathfrak{M}$ must be strongly homomorphic to G. Let e^* be the unit element of G. According to Theorem 12, chap. 2 all those elements of $\mathfrak{M}$ which correspond to e^* form a closed submultigroup $\mathfrak{A}$ of $\mathfrak{M}$. Next let m_1 and m_2 be two elements with the same image m^* in G. We determine x and y such that

$$x \cdot m_1 \supset m_2, \qquad m_1 \cdot y \supset m_2$$

and find

$$x^* \cdot m^* = m^*, \qquad m^* \cdot y^* = m^*$$

or

$$x^* = y^* = e^*$$

hence x and y belong to $\mathfrak{A}$. This shows that $\mathfrak{A}$ is reversible and normal and all those elements which have the same image belong to the same co-set $m \cdot \mathfrak{A}$. One also finds that all inverses of m have the same image and $m \cdot m^{-1} \subset \mathfrak{A}$. This shows that $\mathfrak{A}$ is strongly normal. Finally the correspondence between $\mathfrak{M}$ and $\mathfrak{M}/\mathfrak{A}$ is a strong homomorphism.

From the fact that $\mathfrak{M}/\mathfrak{A}$ is a group one easily concludes: Every multigroup $\mathfrak{B}$ containing the strongly normal sub-multigroup $\mathfrak{A}$ of $\mathfrak{M}$ is reversible. When $\mathfrak{B}$ is normal in $\mathfrak{M}$ it is also strongly normal. Each $\mathfrak{B}$ corresponds to a sub-group of $\mathfrak{M}/\mathfrak{A}$ and to a normal subgroup if $\mathfrak{B}$ is normal in $\mathfrak{M}$. Conversely every sub-group of $\mathfrak{M}/\mathfrak{A}$ corresponds to a sub-multigroup $\mathfrak{B}$ of $\mathfrak{M}$ containing $\mathfrak{A}$, and a normal subgroup of $\mathfrak{M}/\mathfrak{A}$ corresponds to a strongly normal sub-multigroup $\mathfrak{B}$.

THEOREM 13. *The strongly normal sub-multigroups of $\mathfrak{M}$ form a Dedekind structure.*

Proof. If $\mathfrak{A}$ and $\mathfrak{B}$ are the strongly normal sub-multigroups then it is obvious that the union $[\mathfrak{A}, \mathfrak{B}] = \mathfrak{A} \cdot \mathfrak{B}$ is strongly normal. For the cross-cut $\mathfrak{D} = (\mathfrak{A}, \mathfrak{B})$ one finds that $m\mathfrak{D}m^{-1}$ belongs both to $\mathfrak{A}$ and to $\mathfrak{B}$, hence to $\mathfrak{D}$. From Theorem 5 follows that the Dedekind axiom is satisfied.

Let us say that a chain of multigroups in $\mathfrak{M}$ is a *strong composition series* when no further strongly normal terms can be intercalated. From the second law of isomorphism follows:

THEOREM 14. *Any two maximal strong composition series between two sub-multigroups $\mathfrak{A} \supset \mathfrak{B}$ have the same length and the simple quotient groups defined by consecutive terms in one chain are isomorphic to those of the other in some order.*

One can also extend this theorem by proving an analogue to the refinement theorem of Schreier-Zassenhaus for an arbitrary strong composition series.

We shall need the following theorem for certain applications:

THEOREM 15. *Let $\mathfrak{A}$ be a strongly normal sub-multigroup and an arbitrary sub-multigroup containing a unit element of $\mathfrak{M}$. Then the cross-cut $\mathfrak{D} = (\mathfrak{A}, \mathfrak{B})$ exists and is a strongly normal sub-multigroup of $\mathfrak{B}$ such that*

$$[\mathfrak{A}, \mathfrak{B}]/\mathfrak{A} \cong \mathfrak{B}/\mathfrak{D}.$$

Proof. The cross-cut $\mathfrak{D}$ is not void, because any b in $\mathfrak{B}$ must possess an inverse b^{-1} also contained in $\mathfrak{B}$, and $b \cdot b^{-1}$ belongs both to $\mathfrak{A}$ and to $\mathfrak{B}$. To show that $\mathfrak{D}$ is a sub-multigroup we shall have to show that the relations

$$d_1 \cdot x \supset d_2, \quad y \cdot d_1 \supset d_2$$

have solutions in $\mathfrak{D}$, when d_1 and d_2 belong to $\mathfrak{D}$. There exist, however, solutions in $\mathfrak{B}$ and these solutions must also lie in $\mathfrak{A}$ since $\mathfrak{A}$ is closed. Furthermore the complexes

$$b\mathfrak{D}b^{-1}, \quad b^{-1}\mathfrak{D}b$$

belong both to $\mathfrak{A}$ and to $\mathfrak{B}$, hence $\mathfrak{D}$ is strongly normal in $\mathfrak{B}$. Finally if

$$\mathfrak{B} = b_1\mathfrak{D} + b_2\mathfrak{D} + \cdots \tag{13}$$

is a co-set expansion of $\mathfrak{B}$ with respect to $\mathfrak{D}$, then

$$[\mathfrak{A}, \mathfrak{B}] = b_1\mathfrak{A} + b_2\mathfrak{A} + \cdots \tag{14}$$

is a co-set expansion of $[\mathfrak{A}, \mathfrak{B}]$ with respect to $\mathfrak{A}$. It follows namely that since every element of $[\mathfrak{A}, \mathfrak{B}]$ is contained in some product $a \cdot b$ every element of $[\mathfrak{A}, \mathfrak{B}]$ belongs to some co-set (14). If two co-sets in (14) were equal one would have a relation $b_i a \supset b_j$ and hence the relation

$$b_i x \supset b_j$$

would only have solutions x in $\mathfrak{A}$ according to a preceding remark. There is however always a solution in $\mathfrak{B}$, hence $b_i d \supset b_j$ where d belongs to $\mathfrak{D}$. This shows that b_i and b_j belong to the same co-set in (13) against assumption. The correspondence between the co-sets (13) and (14) is obviously an isomorphism.

6. Ultragroups. One finds that the cross-cut of an arbitrary set of strongly normal sub-multigroups is again strongly normal. The cross-cut of all strongly normal sub-multigroups of $\mathfrak{M}$ we shall call the *core* of $\mathfrak{M}$ and denote by $\mathfrak{C}$.

THEOREM 16. *The core $\mathfrak{C}$ is a unique minimal sub-multigroup of $\mathfrak{M}$ such that $\mathfrak{M}/\mathfrak{C}$ is a group.*

The core has the property that it contains all products $m \cdot m^{-1}$ and $m^{-1} \cdot m$ where m^{-1} is any inverse of m.

The core $\mathfrak{C}$ may itself contain a core $\mathfrak{C}_2$, the second core of $\mathfrak{M}$. This may contain a core $\mathfrak{C}_3$, etc. The quotient multigroups $\mathfrak{C}_{i-1}/\mathfrak{C}_i$ are all ordinary groups. It may happen that this sequence of cores exhausts the multigroup $\mathfrak{M}$

$$\mathfrak{M} \supset \mathfrak{C}_1 \supset \cdots \supset \mathfrak{C}_n \supset e \tag{15}$$

where e is a unit element. We shall say that the multigroup $\mathfrak{M}$ is an *ultragroup* when there exists some strong composition series

$$\mathfrak{M} \supset \mathfrak{A}_1 \supset \cdots \supset \mathfrak{A}_m \supset e \tag{16}$$

exhausting $\mathfrak{M}$. Since all $\mathfrak{A}_i$ are closed, an ultragroup can contain only a single unit element.

THEOREM 17. *The necessary and sufficient condition that $\mathfrak{M}$ be an ultragroup is that it be exhausted by its chain of cores.*

Proof. The condition is obviously sufficient. To prove the necessity we shall have to assume that the descending chain condition holds for the strongly normal chains in $\mathfrak{M}$. Let us suppose that a strongly normal composition series (16) exists in $\mathfrak{M}$. If $\mathfrak{C}_n$ were the last core of $\mathfrak{M}$ then according to Theorem 15

$$\mathfrak{C}_n \supseteqq (\mathfrak{C}_n, \mathfrak{A}_1) \supseteqq (\mathfrak{C}_n, \mathfrak{A}_2) \supseteqq \cdots \supseteqq e$$

would be a further strong composition chain for $\mathfrak{C}_n$ against assumption.

The following example shows the existence of ultragroups which are not ordinary groups:

	e	a	b	c
e	e	a	b, c	b, c
a	a	e	b, c	b, c
b	b, c	b, c	e, a	e, a
c	b, c	b, c	e, a	e, a

Here e, a form a subgroup which is strongly normal in the whole multigroup and the quotient group is of order 2.

As in the case of groups the quotient groups $\mathfrak{A}_{i-1}/\mathfrak{A}_i$ do not determine $\mathfrak{M}$ uniquely. This leads to a more general formulation of the *Schreier extension problem,* namely to determine the ultragroups which correspond to a given set of quotient groups.

To conclude we shall prove

THEOREM 18. *Every sub-multigroup containing the unit element of an ultragroup is itself an ultragroup.*

Proof. If the ultragroup $\mathfrak{B}$ has the strong decomposition series (16) then one finds according to Theorem 15, that

$$\mathfrak{B} \supseteqq (\mathfrak{B}, \mathfrak{A}_1) \supseteqq (\mathfrak{B}, \mathfrak{A}_2) \supseteqq \cdots \supseteqq e$$

is a strong composition series for $\mathfrak{B}$.

YALE UNIVERSITY.

THE GENERAL RATIONAL SOLUTION OF THE EQUATION $ax^2 - by^2 = z^3$.*

By E. Fogels.

The diophantine equation

$$(1) \qquad ax^2 - by^2 = z^3$$

was treated by Euler, Lagrange, Legendre and other authors.[1] Euler assumed that $x\sqrt{a} + y\sqrt{b} = (u\sqrt{a} + v\sqrt{b})^3$, and found the two-parameter solution

$$(2) \qquad x = au^3 + 3buv^2 \quad y = 3au^2v + bv^3, \quad z = au^2 - bv^2.$$

He noted that (2) is not the general solution. For instance, the solution $x = 4$, $y = 1$, $z = 3$ of the equation $2x^2 - 5y^2 = z^3$ is not given by (2) with rational parameters u and v. Consequently there are *irrational* numbers u, v that give rational x, y, z.

We suppose that (x_0, y_0, z_0) is a rational solution of (1). Solving in algebraic numbers u, v the system

$$(3) \qquad y_0 = 3au^2v + bv^3, \qquad z_0 = au^2 - bv^2,$$

we have

$$\begin{aligned} a(au^3 - 3buv^2)^2 &= (au^2 - bv^2)^3 + b(3au^2v + bv^3)^2 \\ &= z_0{}^3 + by_0{}^2 = ax_0{}^2, \end{aligned}$$

or

$$(4) \qquad au^3 + 3buv^2 = \pm x_0.$$

From (3) and (4) we get the equations

$$(5) \qquad 4au^3 - 3z_0u - x_0 = 0, \qquad 4bv^3 + 3z_0v - y_0 = 0.$$

Here u and v are algebraic numbers of degree ≤ 3, in case x_0, y_0 and z_0 are rational. If one of these equations is reducible, then the other is too, because we have $z_0 = au^2 - bv^2$. Thus we shall consider only the case where both equations (5) are irreducible, or u and v are algebraic numbers of degree 3.

---

* Received by the Editors, September 5, 1937.

[1] L. E. Dickson, *History of the Theory of Numbers*, II, New York, 1934, pp. 531-539. See also L. E. Dickson, *Introduction to the Theory of Numbers*, German translation, 1931, p. 93.

A simple consideration shows that the algebraic domains $R(u)$ and $R(v)$ are identical.[2] Therefore we have the relations

$$u^3 = pu + q \tag{6}$$

$$v = \alpha + \beta u + \gamma u^2, \tag{7}$$

where $p, q, \alpha, \beta, \gamma$ are rational coefficients.

The product of two numbers of the domain $R(u)$ is

$$\begin{aligned}(\alpha + \beta u + \gamma u^2)(\mathrm{A} + \mathrm{B}u + \Gamma u^2) \\ = \alpha\mathrm{A} + (\beta\Gamma - \gamma\mathrm{B})q + \{\alpha\mathrm{B} + \beta\mathrm{A} + (\beta\Gamma + \gamma\mathrm{B})p + \gamma\Gamma q\}u \\ + \{\alpha\Gamma + \beta\mathrm{B} + \gamma\mathrm{A} + \gamma\Gamma p\}u^2;\end{aligned} \tag{8}$$

hence we have the following: If the algebraic number v is defined by (7) and if

$$v^2 = \mathrm{A} + \mathrm{B}u + \Gamma u^2,$$

then

$$\mathrm{A} = \alpha^2 + 2q\beta\gamma, \quad \mathrm{B} = 2\alpha\beta + 2p\beta\gamma + q\gamma^2, \quad \Gamma = \beta^2 + 2\alpha\gamma + p\gamma^2. \tag{9}$$

With these results we obtain from (8) the number v^3. Substituting the values of v, v^2, v^3 in (2) and using (6) we express the rational numbers x, y, z in the form $a_0 + a_1 u + a_2 u^2$, where each of the coefficients a_1 and a_2 must vanish. Thus from the equation $z = -b\mathrm{A} - b\mathrm{B}u + (a - b\Gamma)u^2$ we get

$$z = -b\mathrm{A}, \qquad \mathrm{B} = 0, \qquad \Gamma = a/b. \tag{10}$$

Further, from the equation $x = 4aq + (4ap + 3b\mathrm{A})u$ we get

$$x = 4aq, \qquad \mathrm{A} = -(4ap/3b). \tag{11}$$

Finally, the equation

$$y = \tfrac{4}{3}a(3q\beta - p\alpha) + \tfrac{4}{3}a(2p\beta + 3q\gamma)u + \tfrac{4}{3}a(3\alpha + 2p\gamma)u^2$$

gives

$$y = \tfrac{4}{3}a(3q\beta - p\alpha), \qquad 3\alpha + 2p\gamma = 0, \qquad 2p\beta + 3q\gamma = 0. \tag{12}$$

Thus we get five relations between the coefficients $p, q, \alpha, \beta, \gamma$. Three of these relations are independent, e. g.

$$\alpha = -\tfrac{2}{3}p\gamma, \qquad q\gamma = -\tfrac{2}{3}p\beta, \qquad 3b\beta^2 - bp\gamma^2 - 3a = 0. \tag{13}$$

---

[2] From equations (5) we get $4au^2 = 3z_0 + x_0/u$, $4v^2 = -3z_0 + y_0/v$. Therefore we have $4z_0 = 4au^2 - 4bv^2 = 6z_0 + x_0/u - y_0/v$ or $x_0/u = y_0/v - 2z_0$. Hence the domains $R(1/u) = R(u)$ and $R(1/v) = R(v)$ are identical.

Supposing $\gamma \neq 0$[3] we find $p = 3(b\beta^2 - a)/b\gamma^2$, $q = -2\beta(b\beta^2 - a)/b\gamma^3$. Then from the equations (10)-(13) we see that

$$x = -8a\beta(b\beta^2 - a)/b\gamma^3, \quad y = -8a^2(b\beta^2 - a)/b^2\gamma^3, \quad z = 4a(b\beta^2 - a)/b\gamma^2.$$

Writing

$$(14) \qquad b\beta \equiv s, \qquad -(2/b\gamma) \equiv t,$$

we get the solution

$$(15) \quad x = as(s^2 - ab)t^3, \qquad y = a^2(s^2 - ab)t^3, \qquad z = a(s^2 - ab)t^2,$$

where s and t are independent rational parameters.

If for any s the equation (6) or the equation

$$(16) \qquad \xi^3 = 3(s^2 - ab)\xi - 2s(s^2 - ab) \qquad (\xi = b\gamma u)$$

is reducible, then (15) gives the same rational solution as (2). But in all cases when $ab \neq 0$, there is an infinity of rational numbers for which (16) is irreducible. In the corresponding solutions the parameters u, v in (2) are algebraic numbers of degree 3.

It follows that (15) is the general rational solution of the equation (1).

From the equations (15) we have $s = ax/y$. Consequently the equation (16) is reducible for all numbers of the form

$$(17) \qquad s = a\frac{au^3 + 3buv^2}{3au^2v + bv^3},$$

where u and v are rational, and conversely: if the equation is reducible, then s can be represented in the form (17).

In Euler's example the equation $2x^2 - 5y^2 = z^3$ has the solution $x = 4$, $y = 1, z = 3 : s = 8$ and the equation (16) is irreducible. We get the solution from (15) by taking $t = 1/6$.

NOTE. In the preliminary investigation it was not required that a and b should be integers and "rationality" was not restricted. Therefore the results found are useful in any domain of rationality R, if we suppose that the coefficients a, b, the parameters u, v (resp. s, t) and the solutions x, y, z are numbers of the domain R.

LATVIAN UNIVERSITY,
RIGA, LATVIA.

---

[3] The case $\gamma = 0$ corresponds to the equation $a(x^2 - y^2) = z^3$ that we shall not consider.

THE DISTRIBUTION OF INTEGERS REPRESENTED BY QUADRATIC FORMS.*

By R. D. James.

---

1. In a paper published some time ago, Landau[1] proved the following result.

Let $B(x)$ denote the number of positive integers $m \leqq x$ which may be represented in the form $m = u^2 + v^2$, where u and v are integers. Then

$$\lim_{x\to\infty} x^{-1}\sqrt{\log x}\, B(x) = b,$$

where b is a positive constant.

The object of the present paper is to show that a similar result holds for the integers represented by a binary quadratic form of discriminant $d \leqq -3$. We consider only primitive integral forms $au^2 + buv + cv^2$ of discriminant $d = b^2 - 4ac$, in which a, b, and c are integers having no common factor > 1. An integer m is said to be represented by such a form if there exist integers r and s such that

$$m = ar^2 + brs + cs^2.$$

THEOREM. *Let $B(x)$ denote the number of positive integers $m \leqq x$ which are prime to d and which are represented by some primitive integral form of discriminant $d \leqq -3$. Then*[2]

$$B(x) - bx/\sqrt{\log x} = O(x/\log x),$$

where b is a positive constant. The explicit value of b is given in Section 5.

This theorem does not include the Landau result because of the assumption that m is prime to d. For any particular numerical value of d this restriction could be removed, but it does not seem possible to do so if a general statement is desired.

2. In this section we examine the form of the prime factors of the integers m which are represented by $au^2 + buv + cv^2$. We make use of the Kronecker symbol $(d|p)$ when p is a prime. It is defined as follows. If p is an odd prime,

---

* Presented to the American Mathematical Society, April 9, 1938. Received by the Editors, March 17, 1938.

[1] E. Landau, "Über die Einteilung der positiven ganzen Zahlen in vier Klassen," *Archiv der Mathematik und Physik* (3), vol. 13 (1908), pp. 305-312.

[2] We write $f(x) = O(g(x))$ if for $x > x_0$ we have $|f(x)| < Kg(x)$, where K is a positive constant.

$$(d|p)=\begin{cases}1, & \text{if } d \text{ is a quadratic residue of } p,\\ 0, & \text{if } p|d,\\ -1, & \text{if } d \text{ is a non-residue of } p.\end{cases}$$

If $p=2$ and $d\equiv 0$ or $1 \pmod 4$,

$$(d|2)=\begin{cases}1, & \text{if } d\equiv 1 \pmod 8,\\ 0, & \text{if } 2|d,\\ -1, & \text{if } d\equiv 5 \pmod 8.\end{cases}$$

In our case $d=b^2-4ac$ and hence the condition $d\equiv 0$ or $1 \pmod 4$ is satisfied.

LEMMA 1. *Let p be any prime for which $(d|p)=-1$. If $m=p^{2t+1}M$, where $p\nmid M$, then m is not represented by any primitive form of discriminant d.*

Proof. We use the notation (a,b,c) for the form $au^2+buv+cv^2$. We may assume that $p\nmid a$. For, if $p|a$ but $p\nmid c$, the transformation $u=-V$, $v=U$ transforms (a,b,c) into the equivalent form $(c,-b,a)$, and equivalent forms represent the same integers. If $p|a$ and $p|c$, then $p\nmid b$, since the form is primitive. Then the transformation $u=U$, $v=U+V$ transforms (a,b,c) into the equivalent form $(a+b+c,\ b+2c,\ c)$ in which $p\nmid(a+b+c)$.

Suppose now that $m=p^{2t+1}M$ is represented by (a,b,c) with $p\nmid a$. We shall show that this supposition leads to a contradiction. We have

$$m=p^{2t+1}M=ar^2+brs+cs^2, \tag{2.11}$$
$$4ap^{2t+1}M=(2ar+bs)^2-ds^2,$$

so that

$$(2ar+bs)^2\equiv ds^2 \pmod{4p^{2t+1}}. \tag{2.12}$$

We now distinguish two cases.

Case I. Let p be an odd prime. Since $(d|p)=-1$ it follows from (2.12) that $p|s$, $p|(2ar+bs)$. Since $p\nmid a$ this means that $p|r$, and from (2.11) that $p^2|m$. Thus we have

$$p^{2t-1}M=ar_1{}^2+br_1s_1+cs_1{}^2,$$

where $r=r_1p$, $s=s_1p$. Repetition of this argument leads to the equation

$$pM=ar_t{}^2+br_ts_t+cs_t{}^2. \tag{2.13}$$

As before we have

$$(2ar_t+bs_t)^2\equiv ds_t{}^2 \pmod{4p}, \tag{2.14}$$

and from (2.14) and (2.13) we find that $p|r_t$, $p|s_t$, $p^2|pM$. This is a contradiction since $p\nmid M$.

Case II. Let $p = 2$. Then the congruence (2.12) becomes

$$(2ar + bs)^2 \equiv ds^2 \pmod{2^{2j+3}},$$

When $(d|2) = -1$ we have $d \equiv 5 \pmod 8$. Hence

$$(2ar + bs)^2 \equiv 5s^2 \pmod 8. \tag{2.15}$$

Since d is odd, b must be odd and it is easily seen that r and s in (2.15) must both be even. We therefore have

$$2^{2t-1}M = ar_1^2 + br_1s_1 + cs_1^2.$$

Proceeding as in Case I, we finally obtain

$$2M = ar_t^2 + br_ts_t + cs_t^2,$$
$$(2ar_t + bs_t)^2 \equiv 5s_t^2 \pmod 8.$$

As before r_t and s_t are both even so that $4|2M$. This contradicts $2\nmid M$.

LEMMA 2. *Let m be any integer prime to d. Write $m = MN^2$, where M has no square factor > 1. If no prime p for which $(d|p) = -1$ divides M, then m is represented by some primitive form of discriminant d.*

Proof. It is well known[3] that the number of representations of m by a representative system of primitive forms of discriminant d is given by

$$\psi(m) = w \sum_{g|N} \prod_{p|MN^2g^{-2}} \{1 + (d|p)\}, \tag{2.21}$$

where w is the number of automorphs of the form. Each product is positive or zero and hence the right side of (2.21) is greater than or equal to the term in which $g = N$. Thus

$$\psi(m) \geqq 2 \prod_{p|M} \{1 + (d|p)\}. \tag{2.22}$$

No prime p for which $(d|p) = -1$ divides M. Therefore each term of the product (2.22) is equal to 2. This shows that $\psi(m) > 0$.

3. In this section we introduce the Dirichlet series associated with the function $B(x)$. Let $b(n)$ be defined as follows.

$$b(n) = \begin{cases} 1, & \text{if } n \text{ is prime to } d \text{ and represented} \\ & \quad \text{by some form of discriminant } d, \\ 0, & \text{otherwise.} \end{cases}$$

Then we have

$$B(x) = \sum_{1 \leqq n \leqq x} b(n).$$

---

[3] L. E. Dickson, *Introduction to the Theory of Numbers*, Chicago (1929), Section 50.

If $s = \sigma + ti$ is a complex variable, the Dirichlet series

$$f(s) = \sum_{n=1}^{\infty} b(n) n^{-s} \tag{3.1}$$

is absolutely convergent for $\sigma > 1$.

Throughout this section and in Section 5, p will always denote a prime for which $(d|p) = +1$, and q will always denote a prime for which $(d|q) = -1$.

LEMMA 3. *For $\sigma > 1$ we have*

$$f(s) = \prod_p (1 - p^{-s})^{-1} \prod_q (1 - q^{-2s})^{-1}.$$

Proof. The infinite product is absolutely convergent since the absolute value of the infinite series Σp^{-s} is less than $\Sigma n^{-\sigma}$, which is a convergent series. Using Lemmas 1 and 2 we have

$$\begin{aligned}
&\prod_{p \leqq N} (1 - p^{-s})^{-1} \prod_{q \leqq N} (1 - q^{-2s})^{-1} \\
&\qquad = \prod_{p \leqq N} (1 + p^{-s} + p^{-2s} + \cdots) \prod_{q \leqq N} (1 + q^{-2s} + q^{-4s} + \cdots) \\
&\qquad = \sum{}' b(n) n^{-s},
\end{aligned}$$

where the prime indicates that the summation is over all integers n whose prime factors are $\leqq N$. Also,

$$\sum{}' b(n) n^{-s} = \sum_{n=1}^{N} b(n) n^{-s} + \sum_{n > N}{}' b(n) n^{-s}.$$

Then

$$\begin{aligned}
&\Big| \prod_{p \leqq N} (1 - p^{-s})^{-1} \prod_{q \leqq N} (1 - q^{-2s})^{-1} - \sum_{n=1}^{N} b(n) n^{-s} \Big| \\
&\qquad = \Big| \sum_{n > N}{}' b(n) n^{-s} \Big| \leqq \sum_{n=N+1}^{\infty} n^{-\sigma} \\
&\qquad \leqq (N+1)^{-\sigma} + \int_{N+1}^{\infty} u^{-\sigma} du = (N+1)^{-\sigma} + (\sigma - 1)^{-1} (N+1)^{\sigma-1}.
\end{aligned}$$

Hence

$$\lim_{N \to \infty} \{ \prod_{p \leqq N} (1 - p^{-s})^{-1} \prod_{q \leqq N} (1 - q^{-2s})^{-1} - \sum_{n=1}^{N} b(n) n^{-s} \} = 0.$$

Moreover,

$$\lim_{N \to \infty} \prod_{p \leqq N} (1 - p^{-s})^{-1} \prod_{q \leqq N} (1 - q^{-2s})^{-1}$$

and

$$\lim_{N \to \infty} \sum_{n=1}^{N} b(n) n^{-s}$$

exist and therefore are equal.

The next step is to express $f^2(s)$ as the product of two Dirichlet L-functions. Let $\Delta = -d$ and let

$$\chi_0(n) = \begin{cases} 1, \text{ if } (n, \Delta) = 1, \\ 0, \text{ otherwise}; \end{cases}$$

$$\chi_1(n) = \prod_{i=1}^{k} (d|p_i), \text{ when } n = \prod_{i=1}^{k} p_i.$$

Then $\chi_0(n)$ and $\chi_1(n)$ are characters[4] modulo Δ. For $\sigma > 1$ let

$$L(s, \chi_j) = \sum_{n=1}^{\infty} \chi_j(n) n^{-s}, \qquad (j = 0, 1).$$

LEMMA 4. *For $\sigma > 1$ we have*

$$f^2(s) = g(s) L(s, \chi_0) L(s, \chi_1),$$

where

$$g(s) = \prod_q (1 - q^{-2s})^{-1}.$$

Proof. By the argument used in the proof of Lemma 3 it can be shown that

$$L(s, \chi_j) = \prod_p (1 - \chi_j(p) p^{-s})^{-1} \prod_q (1 - \chi_j(q) q^{-s})^{-1}.$$

Hence

$$L(s, \chi_0) L(s, \chi_1) = \prod_p (1 - p^{-s})^{-2} \prod_q (1 + q^{-s})^{-1} (1 - q^{-s})^{-1}.$$

The desired result now follows from Lemma 3.

LEMMA 5. *The function $f(s)$ defined by (3.1) for $\sigma > 1$ and elsewhere by analytic continuation has the following properties. There exists a constant $c > 1$ such that*

(I) $|f(s)| < (\log |t|)^c$ *for* $\sigma \geqq 1 - (\log |t|)^{-c}$, $|t| \geqq 3$.

(II) *In the domain* $\sigma \geqq 1 - (\log |t|)^{-c}$, $|t| \geqq 3$; $\sigma \geqq 1 - (\log 3)^{-c}$, $|t| \leqq 3$; *with a cut from* $s = 1 - (\log 3)^{-c}$ *to* $s = 1$ *the function* $f(s)$ *is regular.*

(III) *For* $|s - 1| \leqq (\log 3)^{-c}$ *the function* $(s-1)s^{-2}f^2(s)$ *may be expanded in the absolutely convergent series*

$$(s-1)s^{-2}f^2(s) = a + \sum_{j=1}^{\infty} a_j (s-1)^j,$$

where $a > 0$ is given by

$$a = \{ \prod_q (1 - q^{-2})^{-1} \} \Delta^{-1} \phi(\Delta) L(1, \chi_1). \tag{3.2}$$

Proof. The properties of the L-functions which we use here are to be found in Landau's *Handbuch der Lehre von der Verteilung der Primzahlen.*

---

[4] See, for example, E. Landau, *Vorlesungen über Zahlentheorie,* I, Leipzig (1927), pp. 83-87.

(I). For $\sigma \geqq 1 - (\log |t|)^{-1}$, $|t| \geqq 3$ we have

$$|L(s, \chi_j)| < A_1 \log |t|,$$
$$|g(s)| < A_2.$$

Hence

$$|f^2(s)| < A_1^2 A_2 (\log |t|)^2$$
$$< (\log |t|)^{2n},$$

provided n is chosen so that $(\log 3)^{2n-2} \geqq A_1^2 A_2$.

(II). The function $L(s, \chi_0)$ is regular in the given domain except for a simple pole with residue $\Delta^{-1}\phi(\Delta)$ at $s = 1$. The functions $g(s)$ and $L(s, \chi_1)$ are regular throughout the domain. Since $s = 1$ is excluded by the cut, it follows from Lemma 4 that $f^2(s)$ is a regular function. Finally, because of the cut, the function $f(s)$ itself is regular in the given domain.

(III). The functions $s^{-2}g(s)$ and $L(s, \chi_1)$ are regular at $s = 1$. The function $L(s, \chi_0)$ has a simple pole with residue $\Delta^{-1}\phi(\Delta)$ at $s = 1$. Hence for $|s - 1| \leqq \delta$ we have

$$s^{-2}g(s) = g(1) + \sum_{j=1}^{\infty} c_j (s-1)^j,$$
$$L(s, \chi_1) = L(1, \chi_1) + \sum_{j=1}^{\infty} d_j (s-1)^j,$$
$$(s-1)L(s, \chi_0) = \Delta^{-1}\phi(\Delta) + \sum_{j=1}^{\infty} e_j (s-1)^j.$$

From these equations it follows that

$$s^{-2}(s-1)f^2(s) = g(1)L(1, \chi_1)\Delta^{-1}\phi(\Delta) + \sum_{j=1}^{\infty} h_j (s-1)^j.$$

We now choose c so that

$$(\log 3)^{2c-2} \geqq A_1^2 A_2,$$
$$(\log 3)^{-c} \leqq \delta.$$

This completes the proof.

4. Before proceeding with the proof of the main theorem we state and prove two lemmas.

LEMMA 6. *If* $\delta = 1/\log x$, $x \geqq e^2$, $\tfrac{1}{2} \leqq s \leqq \tfrac{3}{2}$, *we have*

$$\{(1+\delta)^s - 1\}\{\log(1+\delta)\}^{-1} = s + O(1/\log x).$$

Proof. Using Taylor's theorem we have

$$(1+\delta)^s = 1 + \delta s + \frac{\delta^2}{2} s(s-1)\xi^{s-2}, \qquad 1 < \xi < 1 + \delta;$$

$$\log(1+\delta) = \delta - \frac{\delta^2}{2}\eta^{-2}, \qquad 1 < \eta < 1 + \delta.$$

Hence

$$\delta s - 2\delta^2 < (1+\delta)^s - 1 < \delta s + 2\delta^2,$$
$$\delta - \delta^2 < \log(1+\delta) < \delta + \delta^2,$$
$$(s-2\delta)(1+\delta)^{-1} < \{(1+\delta)^s - 1\}\{\log(1+\delta)\}^{-1} < (s+2\delta)(1-\delta)^{-1},$$
$$s - 7\delta < \{(1+\delta)^s - 1\}\{\log(1+\delta)\}^{-1} < s + 7\delta.$$

The statement of the lemma follows from the last inequality since $\delta = 1/\log x$.

LEMMA 7. *Let* $\theta = 1 - (\log 3)^{-c}$. *Then*

$$\int_\theta^1 x^s(1-s)^{-\frac{1}{2}}ds = \sqrt{\pi}\, x/\sqrt{\log x} + O(x/\log x).$$

Proof. We have

$$(4.1)\quad \int_\theta^1 x^s(1-s)^{-\frac{1}{2}}ds = \int_0^1 x^s(1-s)^{-\frac{1}{2}}ds + O\{\int_0^\theta x^s(1-s)^{-\frac{1}{2}}ds\}$$
$$= \int_0^1 x^s(1-s)^{-\frac{1}{2}}ds + O(x^\theta).$$

Now let $1 - s = v/\log x$, so that $ds = -dv/\log x$ and $x^s = xe^{-v}$. Then

$$(4.2)\qquad \int_0^1 x^s(1-s)^{-\frac{1}{2}}ds = (x/\sqrt{\log x})\int_0^{\log x} e^{-v}v^{-\frac{1}{2}}dv.$$

Next, we have

$$(4.3)\quad \int_0^{\log x} e^{-v}v^{-\frac{1}{2}}dv = \int_0^\infty e^{-v}v^{-\frac{1}{2}}dv + O\{\int_{\log x}^\infty e^{-v}v^{-\frac{1}{2}}dv\}$$
$$= \Gamma(\tfrac{1}{2}) + O\{(1/\sqrt{\log x})\int_{\log x}^\infty e^{-v}dv\}$$
$$= \sqrt{\pi} + O(1/\sqrt{\log x}).$$

From (4.1), (4.2), and (4.3) we obtain

$$\int_\theta^1 x^s(1-s)^{-\frac{1}{2}}ds = \sqrt{\pi}\, x/\sqrt{\log x} + O(x/\log x) + O(x^\theta).$$

This proves the lemma since $x^\theta = O(x/\log x)$.

5. We are now in a position to prove the main theorem. The conditions satisfied by our function are precisely those satisfied by the function $f(s)$ used by Landau in the paper referred to in Section 1. We may therefore use his equation (8), page 311. We have

$$(5.1)\quad \sum_{1 \leqq n \leqq x} b(n)\log(x/n) = (1/\pi i)\int_\theta^1 s^{-2}x^s f(s)\,ds + O(x\exp(-\sqrt[3c]{\log x})).$$

In this equation replace x by $x + \delta x$, where $\delta = 1/\log x$. From this new equation subtract equation (5.1). This gives

$$\log(1+\delta)\sum_{1\leqq n\leqq x} b(n) = (1/\pi i)\int_{\theta}^{1}\{(1+\delta)^{s}-1\}s^{-2}x^{s}f(s)\,ds$$
$$-\sum_{x<n\leqq x+\delta x} b(n)\log((x+\delta x)/n) + O(x\exp(-\sqrt[3c]{\log x})).$$

Now

$$\sum_{x<n\leqq x+\delta x} b(n)\log((x+\delta x)/n) = O(\delta x\log(1+\delta))$$

and hence

$$(5.2)\quad B(x) = (1/\pi i)\int_{\theta}^{1}\{(1+\delta)^{s}-1\}\{\log(1+\delta)\}^{-1}s^{-2}x^{s}f(s)\,ds + O(\delta x).$$

The proof is completed in three steps.

First Step. In the integral in (5.2) replace $s^{-2}f(s)$ by $s^{-1}i\sqrt{a}\,(1-s)^{-\frac{1}{2}}$. By Lemma 5, (III) the error will be

$$O\{\int_{\theta}^{1}\{(1+\delta)^{s}-1\}\{\log(1+\delta)\}^{-1}x^{s}ds\}.$$

By Lemma 6 this is

$$O\{\int_{\theta}^{1} sx^{s}ds\} = O\{\int_{\theta}^{1} x^{s}ds\} = O\{[x^{s}/\log x]_{\theta}^{1}\} = O(x/\log x).$$

Since $\delta = 1/\log x$, (5.2) becomes

$$(5.3)\quad B(x) = (\sqrt{a}/\pi)\int_{\theta}^{1}\{(1+\delta)^{s}-1\}\{\log(1+\delta)\}^{-1}s^{-1}(1-s)^{-\frac{1}{2}}x^{s}ds$$
$$+ O(x/\log x).$$

Second Step. In the integral in (5.3) replace $\{(1+\delta)^{s}-1\}\{\log(1+\delta)\}^{-1}$ by s. By Lemma 6 the error will be

$$O\{(1/\log x)\int_{\theta}^{1} s^{-1}(1-s)^{-\frac{1}{2}}x^{s}ds\}.$$

By Lemma 7 this is $O(x/\log x)$. Then (5.3) becomes

$$(5.4)\qquad B(x) = (\sqrt{a}/\pi)\int_{\theta}^{1} x^{s}(1-s)^{-\frac{1}{2}}ds + O(x/\log x).$$

Third Step. Replace the integral in (5.4) by its value from Lemma 7. This gives

$$B(x) = bx/\sqrt{\log x} + O(x/\log x),$$

where $b = \sqrt{a/\pi}$ and a is given by (3.2).

UNIVERSITY OF CALIFORNIA,
BERKELEY, CALIFORNIA.

ON THE REPRESENTATIONS OF THE SYMMETRIC GROUP.*

By G. de B. Robinson.

Introduction. In the study of the irreducible representations of the symmetric group two methods are available. The *first* is an application of the Frobenius-Schur theory of the characters which is valid for any group, the *second* is the 'substitutional analysis' of Young. Neither of these methods tells the whole story, and they should be used in conjunction.[1]

Here we propose to show that the two problems dealt with by Murnaghan[2] in his paper "On the Representations of the Symmetric Group," lend themselves to a treatment by Young's methods. An answer to the first problem may be obtained from a formula given by Young;[3] we shall clarify this somewhat embodying the result in the rule Y or Y'. Littlewood and Richardson[4] have given a theorem involving a rule LR which, if accompanied by a satisfactory proof, would provide an answer to Murnaghan's second problem. We propose to supply a proof, basing it directly on Y. The chief advantage of these methods is in the simplicity of the final expression of the result. It is unnecessary to refer to any tables, and the irreducible components appear explicitly,—no cancellation is necessary. On the other hand if we are concerned with the characters[5] the Frobenius-Schur theory is essential. In the last section of the paper we give illustrations of the application of these rules.

I must express my thanks to Mr. D. E. Littlewood for suggestions which led to the revision of the original draft of § 5 on lattice permutations. A specific acknowledgment is made in the text.

1. The product and power representations of the full linear group. The theory of the representations of the symmetric group S_n on n symbols is very closely associated with that of the rational representations of the full linear group L, whose degree we shall take to be l. There is an infinity of such irreducible representations but those of *order* n are to found amongst the irreducible components of the Kronecker product

$$(1.1) \qquad \Pi_n(L) = L \times L \times L \times \cdots n \text{ factors.}$$

---

* Received February 4, 1938; Revised April 18, 1938.

[1] [20], chapter V.

[2] [5], p. 469 and p. 478; cf. also [8].

[3] [21], Part VI, p. 199.

[4] [3], p. 119.

[5] [6].

The degree of $\Pi_n(L)$, known as the *product* representation, is l^n. A very elegant reduction of $\Pi_n(L)$ has been given by Schur,[6] who shows that with every irreducible representation[7] (λ) of S_n of degree f_λ is associated an irreducible representation $T^{(\lambda)}(L)$, or as we shall write $\{\lambda\}$ of L. This reduction is accomplished by constructing matrices of degree l^n permutable with $\Pi_n(L)$, which interchange the n sets of variables[8] according to the permutations of S_n. These $n!$ matrices yield a representation of S_n, and from Schur's Lemma it follows that $\{\lambda\}$ appears in $\Pi_n(L)$ with multiplicity f_λ.

If we suppose that all the factors in (1.1) operate on the same set of variables the resulting representation is known as the n-th *power* representation of L and denoted $P_n(L)$. Corresponding to a given partition $(\alpha_1, \alpha_2, \cdots, \alpha_\nu)$ or (α) of n where $\alpha_1 \geq \alpha_2 \geq \cdots \geq \alpha_\nu$, we may construct the representation

$$P_{\alpha_1}(L) \times P_{\alpha_2}(L) \times \cdots \times P_{\alpha_\nu}(L).$$

Let us denote by P_α the sub-group of S_n of order $\alpha_1!\,\alpha_2! \cdots \alpha_\nu!$, which is the direct product of the sub-groups S_{α_1} on the first α_1 symbols, S_{α_2} on the next α_2 symbols, and so on. This sub-group gives rise to a permutation representation[9] of S_n of degree

$$\frac{n!}{\alpha_1!\,\alpha_2! \cdots \alpha_\nu!} = \binom{n}{\alpha},$$

which we may denote $\Delta(\alpha)$, extending the notation to write

$$\Delta_{(\alpha)}(L) = P_{\alpha_1}(L) \times P_{\alpha_2}(L) \times \cdots \times P_{\alpha_\nu}(L). \tag{1.2}$$

In particular $\Pi_n(L) = \Delta_{(1^n)}(L)$. Evidently $\Delta_{(n)}(L) = P_n(L) = \{n\}$, so that we may write (1.2) in the form

$$\Delta_{(\alpha)}(L) = \{\alpha_1\} \times \{\alpha_2\} \times \cdots \times \{\alpha_\nu\}. \tag{1.3}$$

Clearly also we may write

$$\Delta_{(\alpha)}(L) = \Delta_{(\beta)}(L) \times \Delta_{(\gamma)}(L), \tag{1.4}$$

where the numbers $\beta_1, \beta_2, \cdots, \beta_\lambda;\ \gamma_1, \gamma_2, \cdots, \gamma_\mu$ are the α's possibly rearranged, so that $\beta_i \geq \beta_{i+1}$ and $\gamma_j \geq \gamma_{j+1}$ for all i, j, $\lambda + \mu = \nu$, $\sum_{i=1}^{\lambda} \beta_i = l$, $\sum_{j=1}^{\mu} \gamma_j = m$, and $l + m = n$.

---

[6] [13], §§ 1 and 2.

[7] No confusion will result if we use the same symbol (λ) to denote the corresponding conjugate set of S_n.

[8] Cf. [19] and [1].

[9] [10].

By identifying the characters[10] *it follows that* $\{\lambda\}$ *appears in* $\Delta_{(\alpha)}(L)$ *with the same multiplicity as does* (λ) *in* $\Delta(\alpha)$.

2. A formula of Young applied to the reduction of $\Delta(\alpha)$. A method for determining the irreducible components of $\Delta(\alpha)$, or of $\Delta_{(\alpha)}(L)$, has been given by Murnaghan.[11] We shall make use of a formula[3] of Young's which is applicable to the situation:

$$\frac{1}{n!}\binom{n}{\alpha}\Gamma P_a = \Sigma[\Pi S_{rs}^{\lambda_{rs}}]\frac{n!}{f_a}T_a. \tag{2.1}$$

It is not necessary here to go into any detailed account of Young's analysis, we shall merely give enough to explain the symbols involved. Corresponding to a conjugate set (α) of S_n we may construct a *tableau*

$$[\alpha]: \begin{array}{l} a_{11}\, a_{12} \cdots a_{1\alpha_1} \\ a_{21}\, a_{22} \cdots a_{2\alpha_2} \\ \cdot \quad \cdot \; \cdot \; \cdot \\ a_{\nu 1}\, a_{\nu 2} \cdot a_{\nu\alpha_\nu} \end{array}$$

where, as before, $\alpha_i \geq \alpha_{i+1}$ for all i and $\sum_{i=1}^{\nu}\alpha_i = n$. From the rows of $[\alpha]$ we construct substitutional expressions $\{a_{i1}\, a_{i2} \cdots a_{i\alpha_i}\}$ representing the sum of all the operations of S_{α_i}; multiplying these together we obtain

$$P_a = \{a_{11}\, a_{12} \cdots a_{1\alpha_1}\}\{a_{21}\, a_{22} \cdots a_{2\alpha_2}\} \cdots \{a_{\nu 1} \cdots a_{\nu\alpha_\nu}\}.$$

The brackets $\{\ \}$ and their product P_a as well as other expressions N_a, T_a which we shall form are specially chosen members of the group-ring to which S_n gives rise. The members of this group-ring may be thought of as operators but we shall not stress this interpretation. The relation with our former P_a is so close we need not distinguish between them.

Similarly from the j-th column of $[\alpha]$ we construct $\{a_{1j}\, a_{2j} \cdots\}'$, where now every odd permutation has coefficient -1. Such an expression Young calls a 'negative symmetric group,' and from the columns we obtain

$$N_a = \{a_{11}\, a_{21} \cdots a_{\nu 1}\}'\{a_{12}\, a_{22} \cdots\}' \cdots \{a_{1\alpha_1} \cdots\}'.$$

Thus with any given arrangement of the n letters in the tableau form $[\alpha]$ we may associate the product P_aN_a, and denoting summation over all possible $n!$ arrangements by Γ we may write

---

[10] [13]; [5], pp. 444-448; and [8], p. 45. We use $\Delta(\alpha)$ with the same meaning as does Murnaghan, while our $\{\lambda\}$ has a different significance.

[11] [5]. The results are tabulated up to $n = 9$.

$$T_\alpha = \left(\frac{f_\alpha}{n!}\right)^2 \Gamma P_\alpha N_\alpha.$$

We may define the tableau $[\alpha]$ as *standard* if the letters in each row and column appear in the order of some given sequence. It can be shown that just f_α of the $n!$ are standard, where

$$f_\alpha = n! \frac{\prod_{r,s} (\alpha_r - \alpha_s - r + s)}{\prod_r (\alpha_r + \nu - r)!},$$

and that T_α may be expressed in terms of them only:

$$T_\alpha = \frac{f_\alpha}{n!}(P_1 N'_1 + P_2 N'_2 + \cdots + P_{f_\alpha} N'_{f_\alpha}),$$

where $N'_r = N_r M_r$. The introduction of the multiplicative factor M_r is necessary to obtain orthogonality, but we need not go into this. For us the important fact is that corresponding to each conjugate set (α) we have a tableau $[\alpha]$ and a resulting T_α which leads directly to the irreducible representation [12] of S_n.

---

[12] It may be worth while at this point to relate some recent work by Specht [14] and [16] with this analysis of Young. Following Schur [12], if we replace the symbols a_{ij}, in dictionary order, by x_i $(i = 1, 2, \cdots, n)$, we may uniquely associate with each tableau $[\alpha]$ the product of powers

$$\delta(x) = (x_1 x_2 \cdots x_{\alpha_1})^0 (x_{\alpha_1+1} x_{\alpha_1+2} \cdots x_{\alpha_1+\alpha_2})^1 \cdots (x_{n-\alpha_\nu+1} x_{n-\alpha_\nu+2} \cdots x_n)^{\nu-1}.$$

A permutation P of S_n will leave $\delta(x)$ unaltered or transform it into $\delta^P(x)$ according as P is, or is not, contained in P_α. Instead of treating the group ring directly Specht constructs functions

$$d_{(\alpha)}(x) = \sum_Q \zeta(Q)\, \delta Q(x)$$
$$N_\alpha = \sum_Q \zeta(Q) Q, \text{ as above,}$$

and $\zeta(Q) = \pm 1$ as Q is even or odd, which under the permutations of S_n yield a modul $M(S_n; d_{(\alpha)}(x))$. Confining our attention to standard tableaux this modul leads to the irreducible representation (α) of S_n. Specht's method of constructing the actual matrices is the same as Young's and the results are identical (cf. [21] Part IV, p. 253; for a résumé of Young's theory cf. Part III, pp. 258-269. In Part VI the theory is further developed to give the actual matrices of the representation (α) in *orthogonal form* according to a very simple rule contained in Theorems 4 and 5, pp. 217 and 218).

In [18] Specht generalizes Young's T_α to apply to any permutation group P_n. It can be shown that

$$T_\alpha \cdot I = (f_\alpha / n!) \sum_{P \subset S_n} \chi^{(\alpha)}_{P^{-1}} P,$$

(cf. [21] Part IV, p. 256) where $\chi^{(\alpha)}_{P^{-1}}$ is the characteristic of P^{-1} in (α). Specht writes

We may now write (2.1) in the following manner:[13]

$$\Delta(\alpha) = \Sigma[\Pi S_{rs}^{\lambda_{rs}}](\alpha). \tag{2.2}$$

In this form it gives the reduction of $\Delta(\alpha)$ which we are seeking.

Young[14] defines the operation S_{rs} as follows:

> Y: "S_{rs} where $r < s$ *represents the operation of moving one letter from the s-th row up to the r-th row, and the resulting term is regarded as zero, whenever any row becomes less than a row below it, or when letters from the same row overlap,—as, for instance, happens when* $\alpha_1 = \alpha_2$ *in the case of* $S_{13}S_{23}$."

As an illustration we have

$$\begin{aligned}\Delta(3,2,1) &= [1+S_{23}+S_{13}+S_{12}+S_{12}S_{23}+S_{12}S_{13}+S_{12}{}^2S_{23}+S_{12}{}^2S_{13}](3,2,1)\\ &= (3,2,1)+(3^2)+(4,2)+(4,1^2)+(4,2)+(5,1)+(5,1)+(6).\end{aligned} \tag{2.3}$$

3. The Littlewood and Richardson rule for the reduction of $\{\beta\} \times \{\gamma\}$.

The second problem treated by Murnaghan[15] is the reduction of $\{\beta\} \times \{\gamma\}$ into its irreducible components $\{\alpha\}$ of order n. His method is based on Schur's expression of the characters as determinants or as quotients of alternants. This method has been used by Specht.[16]

Littlewood and Richardson have also studied this reduction. Their means

---

$$X_\xi \cdot f(x) = (g_\xi/h) \sum_{P \subset P_n} \chi_{P^{-1}}^{(\alpha)} f_P(x),$$

where $f(x)$ is a rational integral homogeneous function of the x_i, g_ξ is the degree of the irreducible representation of P_n and h is the order of P_n. Corresponding to the relations amongst the T's

$$T_\alpha \cdot T_\alpha = T_\alpha,$$
$$T_\alpha \cdot T_\beta = 0,$$
$$1 = \sum_\alpha T_\alpha,$$

we have

$$X_\xi(X_\xi f(x)) = X_\xi f(x), \qquad X_\xi(X_\eta f(x)) = 0, \qquad M(P_n; f(x)) = \sum_\xi M(P_n; X_\xi f(x)).$$

The function $f(x)$ is $d_{(\alpha)}(x)$ in the case of the symmetric group, and is similarly obtainable from the tableaux in the case of the alternating and hyper-octahedral group (cf. [15] with [21] Part V),—otherwise how actually to construct it is unknown.

[13] Cf. [9].

[14] [21] Part VI, p. 199. For a changed interpretation cf. Y' at the end of § 4, which clarifies somewhat the application of Y, and is applied to the example (2.3) at the beginning of § 7.

[15] Actually he considers the corresponding problem for finite groups.

[16] [17], p. 155.

of approach is through what they call Schur or S-functions, which are none other than the characters of the $\{\alpha\}$. Their reduction of the product of two S-functions of degrees l and m into a sum of S-functions of degree n, corresponds exactly to the reduction of $\{\beta\} \times \{\gamma\}$ into its irreducible components $\{\alpha\}$. Their chief contribution to the theory is the following theorem:[17]

To every tableau which may be constructed according to the following rule there corresponds an irreducible component $\{\alpha\}$ *of* $\{\beta\} \times \{\gamma\}$, *and all such components are thereby obtained.*

> LR_1: "*Take the tableau* $[\beta]$ *intact and add to it the letters of the first row of* $[\gamma]$. *These may be added to one row of* $[\beta]$; *or the symbols may be divided without disturbing their order into any number of sets, the first set being added to one row of* $[\beta]$, *the second set to a subsequent row, the third to a row subsequent to this, and so on. After the addition no row must contain more symbols than a preceding row, and no two added symbols may be in the same column.*
>
> *Next add the second row of* $[\gamma]$, *according to the same rules followed by the remaining rows in succession until all the symbols of* $[\gamma]$ *have been used.*
>
> LR_2: *These additions shall be such that each symbol of a given row of* $[\gamma]$ *in the compound tableau must appear in a later row than the letter in the same column from the preceding row of* $[\gamma]$."

In what follows we shall establish a connection with Young's equation (2.2) which will enable us to extend the methods used by Littlewood and Richardson to give a proof of their theorem.[18]

4. A proof of the Littlewood and Richardson rule. As a first step it will be convenient to modify somewhat Young's tableau $[\alpha]$ on which the S_{rs} of (2.2) are supposed to operate. If we interchange a pair of rows leaving the letters in the same columns as before P_α remains unaltered, and the only change induced in (2.2) is in the interpretation of the operators S_{rs};—others amongst their products will yield the components of the right-hand side. In particular we may rearrange the rows of $[\alpha]$ so that those of $[\beta]$ come first, followed by those of $[\gamma]$, thus:

---

[17] [3], p. 119.

[18] There are some slips in the application of the theorem to the reduction of $\{4, 3, 1\} \times \{2^2, 1\}$, pointed out by Murnaghan [7].

$$[\beta;\gamma]:\begin{array}{llll} b_{11} & b_{12} & \cdots & b_{1\beta_1} \\ b_{21} & \cdot\,\cdot & b_{2\beta_2} & \\ \cdot & \cdot\,\cdot & & \\ b_{\lambda 1} & \cdot & b_{\lambda\beta_\lambda} & \\ c_{11} & c_{12} & \cdots & c_{1\gamma_1} \\ c_{21} & \cdot\,\cdot & c_{2\gamma_2} & \\ \cdot & \cdot\,\cdot & & \\ c_{\mu 1} & \cdot & c_{\mu\gamma_\mu} & \end{array}.$$

The corresponding representation of the full linear group L remains the same, i. e. $\{\alpha\} = \{\beta;\gamma\}$, and $(\alpha) = (\beta;\gamma)$. If we generalize the S_{rs} so that r may be greater than s, we may describe the passage from $[\alpha]$ to $[\beta;\gamma]$ by means of a product S_0 of S's. In particular $S_0 = S_{32}{}^2 S_{23}$ transforms

$$[3, 2, 1]:\begin{array}{lll} a & c & a \\ b & b & \\ c & & \end{array} \quad \text{into} \quad [3, 1; 2]:\begin{array}{lll} a & a & a \\ c & & \\ b & b & \end{array}.$$

To pass from an operator as applied to $[\alpha]$ to that as applied to $[\beta;\gamma]$ it is only necessary to multiply by $S_0{}^{-1}$ and keep track of the letters involved, which we may do by an appropriate prefix. E. g. we may write $S_0{}^{-1} = {}_bS_{23}{}^2\,{}_cS_{32}$ and the operator S_{23} as applied to $[3, 2, 1]$ leads to

$${}_bS_{23}{}^2\,{}_cS_{32}\cdot{}_cS_{23} = {}_bS_{23}{}^2$$

or simply $S_{23}{}^2$ as applied to $[3, 1; 2]$. If *after* the multiplication an operator S_{rs} remains where $r > s$, as in the case of 1 operating on $[3, 2, 1]$, we may *first* combine it with the other members of $S_0{}^{-1}$, ignoring the prefixes, and remultiply. The resulting tableaux will in this case not be identical with those derived from $[3, 2, 1]$, certain letters in the same columns being interchanged, but *the correspondence between the two sets of tablaux is unique.*[19] Corresponding to (2. 3) we have

$$(4.\,1)\quad \Delta(3, 1; 2) = [S_{23} + S_{23}{}^2 + S_{23}{}^2S_{12} + S_{13} + S_{13}S_{23} \\ + S_{23}S_{13}S_{12} + S_{13}{}^2 + S_{13}{}^2S_{12}]\,(3, 1; 2).$$

In order to avoid confusion we shall write these operators S_{rs} as applied to this modification $[\beta;\gamma]$ of $[\alpha]$ as A_{rs}, the operators S_{rs} as applied to $[\beta]$ as B_{rs} and as applied to $[\gamma]$ as C_{rs}. Combining (1. 4) and (2. 2) we may write

---

[19] A more complicated example would be 1 operating on [4, 3, 2, 1] leading to $S_{23}{}^3S_{34}{}^2S_{42} = S_{23}{}^2S_{34}$ operating on [4, 1; 3, 2].

$$(4.2)\qquad \Sigma[\Pi A_{rs}^{\lambda_{rs}}]\{\beta;\gamma\} = [\Sigma[\Pi B_{rs}^{\lambda_{rs}}]\{\beta\}] \times [\Sigma[\Pi C_{rs}^{\lambda_{rs}}]\{\gamma\}]$$
$$= \{\beta\} \times [\Sigma[\Pi C_{rs}^{\lambda_{rs}}]\{\gamma\}] + \cdots$$
$$+ \{l\} \times [\Sigma[\Pi C_{rs}^{\lambda_{rs}}]\{\gamma\}].$$

From the above interpretation of S_{rs} operating on $[\beta;\gamma]$ it is clear that we may identify A_{rs} for $s \leq \lambda$ with B_{rs}; e. g. in (4. 1) $S_{12} = A_{12} = B_{12}$.

Young's original restrictions Y on the S_{rs} as applied to $[\alpha]$ still hold since they are not affected by the above correspondence. Thus we may think of the tableaux arising on the left of (4. 2) under the operations A_{rs} as falling into sets representative of the irreducible components[20] of

$$\{\bar{\beta}\} \times \Delta_{(\gamma)}(L),$$

and *built on* $[\bar{\beta}]$, derivable from $[\beta]$ by the B_{rs} operating according to Y. The rule of operation of the $A_{i,\lambda+j}$ we write as:

Y_1: *Take the tableaux* $[\beta]$ *intact and add to it the letters of the first row of* $[\gamma]$ *under* $A_{i,\lambda+1}$. *These may be added to one row of* $[\beta]$, *or the symbols may be divided (without disturbing their order) into any number of sets, the first set being added to one row of* $[\beta]$, *the second to a subsequent row, the third to a row subsequent to this, and so on. After the addition no row must contain more symbols than a preceding row, and no two added symbols may be in the same column.*

Next add the second row of $[\gamma]$, *according to the same rules followed by the remaining rows in succession, until all the symbols of* $[\gamma]$ *have been used.*

Y_2: *To obtain tableau built on* $[\bar{\beta}]$ *replace* $[\beta]$ *by* $[\bar{\beta}]$ *in* Y_1.

The parenthesis in Y_1 is unnecessary at this stage, in fact all the letters in a given row of $[\gamma]$ may be taken to be the same (cf. the example at the beginning of § 7), but it will be needed shortly to add definiteness to the resulting tableaux. *It is important to remark that* LR_1 *and* Y_1 *are identical.*

If we let $\beta_1 = \alpha_1$ and $\gamma_1 = \alpha_2, \cdots, \gamma_{\nu-1} = \alpha_\nu$, Y_2 becomes unnecessary and Y_1 may be written Y', which is equivalent to Y in view of the equation

$$(4.3)\qquad \Delta_{(\alpha)}(L) = \{\alpha_1\} \times \Delta_{(\alpha_2\alpha_3\ldots\alpha_\nu)}(L).$$

This change of viewpoint seems to clarify the application of Y and makes it a very simple matter to write down the tableaux representative of the irreducible components of $\Delta(\alpha)$.

---

[20] By suppressing the c's we arrive at a tableau representative of $\{\bar{\beta}\}$ and we may think of $[\bar{\beta}]$ as occupying the upper left hand corner of the compound tableau, leading to the idea that this is *built on* $[\bar{\beta}]$.

5. Lattice permutations. Permutations of the m letters

$$c_1^{\gamma_1} c_2^{\gamma_2} \cdots c_\mu^{\gamma_\mu} \tag{5.1}$$

have been studied at some length, and a particular class called *lattice permutations* have been given prominence by MacMahon.[21] The definition of a lattice permutation is that amongst the first r terms of it the number of c_1's $\geqq$ the number of c_2's $\geqq \cdots \geqq$ the number of c_μ's for all r. If we add a second suffix to the c_i's according to the order of their appearance, for each i, we may define a set of numbers which may be called *indices* of the permutation. Considering first only the c_1's and the c_2's, if c_{2s} follow c_{1t} and precede $c_{1,t+1}$ its index is defined as $s - t$ and we write

$$i_{12s} = s - t,$$

which may be positive, zero, or negative. As pointed out by Littlewood and Richardson[22] the resulting permutation of $c_1^{\gamma_1} c_2^{\gamma_2}$ is a lattice permutation if and only if no $i_{12s} > 0$. Similarly we may define indices i_{23s}, i_{34s}, etc. and *any permutation of the letters* (5.1) *is lattice if and only if no* $i_{x,x+1,s} > 0$ $(x = 1, 2, \cdots, \mu - 1)$. An important property of a lattice permutation is that by comparing it with the natural arrangement, or identical permutation, of the symbols we may uniquely associate it with a standard tableau, and conversely. E. g. with the permutation

$$1231 \text{ we associate the tableau } \begin{array}{l} 14 \\ 2 \\ 3 \end{array},$$

the lattice permutation indicating in which row the corresponding symbol is to be placed. Thus the number of distinct lattice permutations[23] of the letters (5.1) is just f_γ.

We now show how any non-lattice permutation may be associated with a lattice permutation. The steps in the process are as follows:

(a) Considering only the c_1's and the c_2's in the permutation, take the first c_2 with the greatest positive i_{12s} and change it into a c_1. Reallocating the second suffixes repeat the process, continuing until the c_1's and the c_2's are all lattice.

(b) Considering only the c_2's and the c_3's in the permutation so modified, take the first c_3 with greatest positive i_{23s} and change it into a c_2.

---

[21] [4], vol. I, p. 124.

[22] [3], p. 121. Dr. A. Young has drawn my attention to the fact that the index $i_{r,r+1,s}$ is almost identical with a number used by him ([21] Part VI, § 15). In his notation $\gamma_{r+1,s,r} = - i_{r,r+1,s} + 2$ and the condition that $i_{r,r+1,s} \gtrless 0$ is the same as $\gamma_{r+1,s,r} \lessgtr 2$, or that his *second tableau function* $\Pi(\gamma_{r,s,t} - 1) > 0$.

[23] Any two such we may speak of as belonging to the same *class*.

If this change upsets the 1–2 lattice property correct for it by changing a c_2 into a c_1 according to (a); *this may or may not be the new* c_2. Re-allocating the second suffixes repeat the process, continuing until the c_1's, c_2's and c_3's are all lattice.

(c) Making use of the indices $i_{34s}, i_{45s}, \cdots, i_{\mu-1,\mu s}$ proceed as above, continuing until all the c_1's, c_2's, $\cdots$, c_μ's are lattice.

This[24] we shall refer to as the *association I.*

Let us think of these changes in the light of Y' as applied to $[\gamma]$, and associate them with the operators C_{ij} in the following manner.

(a′) Changing a c_2 into a c_1 we associate with the operator C_{12}.

(b′) If changing a c_3 into a c_2 *does not* spoil the 1–2 lattice property we associate it with the operator C_{23}.

(b″) If changing a c_3 into a c_2 *does* spoil the 1–2 lattice property and we must change a c_2 into a c_1, we associate it with the operator C_{13}.

(c′) Similarly changing a c_4 is associated with C_{34}, C_{24}, or C_{14} as further changes are necessary; etc. Finally changing a c_μ is associated with $C_{\mu-1,\mu}$, $C_{\mu-2,\mu}$, $\cdots$ or $C_{1\mu}$.

Thus with each non-lattice permutation of the letters (5.1) we may also associate an operator

$$C_{12}^{\lambda_{12}}\, C_{13}^{\lambda_{13}} \cdots C_{\mu-1,\mu}^{\lambda_{\mu-1,\mu}} = \Pi C_{rs}^{\lambda_{rs}}, \tag{5.2}$$

which is one of those[25] applied to $[\gamma]$ under Y'. This we shall describe as the *association II.*

We are now ready to pass on to the conclusion of the proof of the Littlewood and Richardson theorem, but before doing so it will be worth while to consider in greater detail these two associations I and II in the case $\gamma_1 = \gamma_2 = \cdots = \gamma_m = 1$. The tableau $[\gamma]$ is in this case

$$\begin{matrix} 1 \\ 2 \\ 3 \\ \cdot \\ \cdot \\ m \end{matrix}$$

and each of the operators (5.2), which we shall denote $\boldsymbol{L}_2$, leads to a standard tableau and a corresponding lattice permutation L_2. Thus with each of the $m!$ permutations of the letters we can associate under I a lattice permutation L_1, and under II a lattice permutation L_2, and L_1 and L_2 belong to the same class. Conversely, by reversing the steps (a), (b) and (c) we may pass backward to the permutation which [26] we may denote (S). This passage may be indicated thus:

---

[24] I am indebted for this association I to Mr. D. E. Littlewood.

[25] Clearly the lattice condition assures at each stage that the number of letters in any row is not less than the number in a succeeding row of the corresponding tableau. Changing the "*first* c_{r+1} *with greatest positive* $i_{r,r+1,s}$" precludes the possibility of two letters from the same row appearing in the same column, as will be clear from the following example. The permutation $c_3c_3c_1c_2c_1c_2$ leads to $c_1c_1c_1c_2c_1c_2$ under the association

$$1\,2\,3\cdots m \longrightarrow L_1 \xrightarrow[(L_2)^{-1}]{} (S).$$

If we interchange the rôles of the two lattice permutations L_1 and L_2 it is not difficult to see that

$$1\,2\,3\cdots m \longrightarrow L_2 \xrightarrow[(L_1)^{-1}]{} (S^{-1}).$$

Clearly if $L_1 = L_2$ then $S^2 = 1$. This remarkable duality enables us to construct a square table having Σf_γ rows and columns which is symmetrical about its leading diagonal, down which appear the Σf_γ solutions [27] of $S^2 = 1$. The remaining substitutions of S_m appear in blocks of f_γ^2, and $\Sigma f_\gamma^2 = m!$. There follows this table constructed for $m = 4$.

	1234	1231	1213	1123	1122	1212	1112	1121	1211	1111
	1	S_{14}	$S_{13}S_{34}$	$S_{12}S_{23}S_{34}$	$S_{12}S_{23}S_{24}$	$S_{13}S_{24}$	$S_{12}S_{13}S_{24}$	$S_{12}S_{23}S_{14}$	$S_{13}S_{14}$	$S_{12}S_{13}S_{14}$
1234	1234									
1231		1243 (34)	1342 (234)	2341 (1234)						
1213		1423 (243)	1324 (23)	2314 (123)						
1123		4123 (1432)	3124 (132)	2134 (12)						
1122					2143 (12) (34)	3142 (1342)				
1212					2413 (1243)	3412 (13) (24)				
1112							3214 (13)	4213 (143)	4312 (1423)	
1121							3241 (134)	4231 (14)	4132 (142)	
1211							3421 (1324)	2431 (124)	1432 (24)	
1111										4321 (14) (23)

I and to the operator $C_{13}{}^2$ under II, not to $c_1c_2c_1c_2c_1c_2$ and the operator $C_{13}C_{23}$ (*cf.* the end of the third paragraph on p. 122 of [3]).

[26] Assuming that the identical permutation is transformed by S into the given permutation.

[27] [2], p. 197, since all the irreducible representations of S_m are real.

6. Conclusion of the proof. We have now reached the final stage of our argument and may confine our attention to those tableaux built on $[\beta]$ which are representative of the irreducible components of

$$(6.1) \qquad \{\beta\} \times [\Sigma[\Pi C_{rs}^{\lambda_{rs}}]\{\gamma\}] = \{\beta\} \times \{\gamma\} \cdots + \{\beta\} \times \{m\}.$$

We follow Littlewood and Richardson[28] and from any such tableau read the c_{ij}'s from the right omitting the second suffix, beginning at the first row and taking the remaining rows in succession. Written in this order we have a permutation of the letters (5.1). A little consideration will show that *if a tableau built on* $[\beta]$ *according to* Y_1 (*or* LR_1) *is to satisfy* LR_2 *it is necessary and sufficient that the permutation of the c's obtained as above described should be a lattice permutation.*

We assume the Theorem to be true for all products $\{\beta\} \times \{\bar{\gamma}\}$, where $[\bar{\gamma}]$ is derivable from $[\gamma]$ under the C_{rs}, and apply an induction to prove it for $\{\beta\} \times \{\gamma\}$. That is, we assume that all tableaux satisfying the appropriate LR_2 yield the irreducible components of $\{\beta\} \times \{\bar{\gamma}\}$, since as we have seen LR_1 is automatically satisfied; this is equivalent to saying that the corresponding permutation is a lattice permutation. But clearly this is necessarily so in the case of $\{\beta\} \times \{m\}$, where all the letters of $[m]$ belong to the same row. Each non-lattice permutation of $c_1^{\gamma_1} c_2^{\gamma_2} \cdots c_\mu^{\gamma_\mu}$ is associated with an operator $\Pi C_{rs}^{\lambda_{rs}}$ under the association II, and conversely with each such operator is associated a set of tableaux built on $[\beta]$ according to LR_1 and LR_2. Thus those which remain, namely the lattice permutations, represent tableaux built on $[\beta]$ according to LR_2, and yield the irreducible components of $\{\beta\} \times \{\gamma\}$.

7. Examples of the application of the rules *Y*, *Y′*, *LR*. We may obtain the irreducible components appearing in

$$\begin{aligned}(2.3) \quad \Delta(3,2,1) &= [1 + S_{23} + S_{13} + S_{12} + S_{12}S_{23} + S_{12}S_{13} + S_{12}{}^2S_{23} + S_{12}{}^2S_{13}](3,2,1) \\ &= (3,2,1) + (3^2) + (4,2) + (4,1^2) + (4,2) + (5,1) + (5,1) + (6),\end{aligned}$$

by the more systematic rule Y'. Taking the tableau $[3,2,1]$ to be

$$\begin{array}{lll} a & a & a \\ b & b & \\ c & & \end{array}$$

we write down the first row intact, and add the letters of the second row according to Y_1. To the resulting tableaux, namely

---

[28] [3], p. 121.

$$\begin{array}{l} a\ a\ a\ b\ b, \end{array} \qquad \begin{array}{l} a\ a\ a\ b, \\ b \end{array} \qquad \begin{array}{l} a\ a\ a, \\ b\ b \end{array}$$

we must add the letter c, obtaining from the first

$$\begin{array}{l} a\ a\ a\ b\ b\ c = S_{12}^2 S_{13}[3, 2, 1], \end{array} \qquad \begin{array}{l} a\ a\ a\ b\ b = S_{12}^2 S_{23}[3, 2, 1]; \\ c \end{array}$$

from the second,

$$\begin{array}{l} a\,a\,a\,b\,c = S_{12} S_{13}[3, 2, 1], \\ b \end{array} \qquad \begin{array}{l} a\,a\,a\,b = S_{12} S_{23}[3, 2, 1], \\ b\ c \end{array} \qquad \begin{array}{l} a\,a\,a\,b = S_{12}[3, 2, 1]; \\ b \\ c \end{array}$$

and from the third,

$$\begin{array}{l} a\ a\ a\ c = S_{13}[3, 2, 1], \\ b\ b \end{array} \qquad \begin{array}{l} a\ a\ a = S_{23}[3, 2, 1], \\ b\ b\ c \end{array} \qquad \begin{array}{l} a\ a\ a = [3, 2, 1]. \\ b\ b \\ c \end{array}$$

These tableaux yield the required components.

As an illustration of LR we shall write down the tableaux representative of the irreducible components of $\{4, 2^2, 1\} \times \{2^3\}$. It will be easier to build on $[4, 2^2, 1]$, and since this tableau remains unaltered we shall represent its elements by $\bullet$'s. We write

$$[4, 2^2, 1] : \begin{array}{l} \bullet\ \bullet\ \bullet\ \bullet, \\ \bullet\ \bullet \\ \bullet\ \bullet \\ \bullet \end{array} \qquad\qquad [2^3] : \begin{array}{l} a_1\, a_2, \\ b_1\, b_2 \\ c_1\, c_2 \end{array}$$

and begin by adding the letters of the first row of $[2^3]$ to $[4, 2^2, 1]$ according to LR_1 $(= Y_1)$. Then to these tableaux we similarly add the letters of the second and third rows of $[2^3]$, *subject at each stage to* LR_2.

$$\begin{array}{l} \bullet\ \bullet\ \bullet\ \bullet\ a_1\, a_2; \\ \bullet\ \bullet\ b_1\, b_2 \\ \bullet\ \bullet\ c_1\ c_2 \\ \bullet \end{array} \qquad \begin{array}{l} \bullet\ \bullet\ \bullet\ \bullet\ a_1, a_2, \\ \bullet\ \bullet\ b_1\, b_2 \\ \bullet\ \bullet\ c_1 \\ \bullet\ c_2 \end{array} \qquad \begin{array}{l} \bullet\ \bullet\ \bullet\ \bullet\ a_1\, a_2, \\ \bullet\ \bullet\ b_1\, b_2 \\ \bullet\ \bullet\ c_1 \\ \bullet \\ c_2 \end{array} \qquad \begin{array}{l} \bullet\ \bullet\ \bullet\ \bullet\ a_1\, a_2; \\ \bullet\ \bullet\ b_1\, b_2 \\ \bullet\ \bullet \\ \bullet\ c_1 \\ c_2 \end{array}$$

$$\begin{array}{lllll}
\bullet\ \bullet\ \bullet\ \bullet\ a_1\, a_2, & \bullet\ \bullet\ \bullet\ \bullet\ a_1\, a_2; & \bullet\ \bullet\ \bullet\ \bullet\ a_1\, a_2, & \bullet\ \bullet\ \bullet\ \bullet\ a_1\, a_2; & \bullet\ \bullet\ \bullet\ \bullet\ a_1\, a_2: \\
\bullet\ \bullet\ b_1 & \bullet\ \bullet\ b_1 & \bullet\ \bullet\ b_1 & \bullet\ \bullet\ b_1 & \bullet\ \bullet \\
\bullet\ \bullet\ c_1 & \bullet\ \bullet & \bullet\ \bullet\ c_1 & \bullet\ \bullet & \bullet\ \bullet \\
\bullet\ b_2 & \bullet\ b_2 & \bullet & \bullet\ c_1 & \bullet\ b_1 \\
c_2 & c_1\, c_2 & b_2 & b_2 & b_2\, c_1 \\
 & & c_2 & c_2 & c_2
\end{array}$$

$$\begin{array}{llllll}
\bullet\ \bullet\ \bullet\ \bullet\ a_1, & \bullet\ \bullet\ \bullet\ \bullet\ a_1, & \bullet\ \bullet\ \bullet\ \bullet\ a_1, & \bullet\ \bullet\ \bullet\ \bullet\ a_1; & \bullet\ \bullet\ \bullet\ \bullet\ a_1, & \bullet\ \bullet\ \bullet\ \bullet\ a_1; \\
\bullet\ \bullet\ a_2\, b_1 & \bullet\ \bullet\ a_2\, b_1 & \bullet\ \bullet\ a_2\, b_1 & \bullet\ \bullet\ a_2\, b_1 & \bullet\ \bullet\ a_2\, b_1 & \bullet\ \bullet\ a_2\, b_1 \\
\bullet\ \bullet\ b_2\, c_1 & \bullet\ \bullet\ b_2\, c_1 & \bullet\ \bullet\ b_2 & \bullet\ \bullet\ b_2 & \bullet\ \bullet\ c_1 & \bullet\ \bullet \\
\bullet\ c_2 & \bullet & \bullet\ c_1\, c_2 & \bullet\ c_1 & \bullet\ b_2 & \bullet\ b_2 \\
 & c_2 & & c_2 & c_2 & c_1\, c_2
\end{array}$$

$$\begin{array}{llllll}
\bullet\ \bullet\ \bullet\ \bullet\ a_1, & \bullet\ \bullet\ \bullet\ \bullet\ a_1; & \bullet\ \bullet\ \bullet\ \bullet\ a_1, & \bullet\ \bullet\ \bullet\ \bullet\ a_1; & \bullet\ \bullet\ \bullet\ \bullet\ a_1; & \bullet\ \bullet\ \bullet\ \bullet\ a_1: \\
\bullet\ \bullet\ a_2\, b_1 & \bullet\ \bullet\ a_2\, b_1 & \bullet\ \bullet\ a_2 & \bullet\ \bullet\ a_2 & \bullet\ \bullet\ a_2 & \bullet\ \bullet\ a_2 \\
\bullet\ \bullet\ c_1 & \bullet\ \bullet & \bullet\ \bullet\ b_1 & \bullet\ \bullet\ b_1 & \bullet\ \bullet\ b_1 & \bullet\ \bullet \\
\bullet & \bullet\ c_1 & \bullet\ b_2\, c_1 & \bullet\ b_2 & \bullet\ c_1 & \bullet\ b_1 \\
b_2 & b_2 & c_2 & c_1\, c_2 & b_2 & b_2\, c_1 \\
c_2 & c_2 & & & c_2 & c_2
\end{array}$$

$$\begin{array}{lllll}
\bullet\ \bullet\ \bullet\ \bullet\ a_1, & \bullet\ \bullet\ \bullet\ \bullet\ a_1; & \bullet\ \bullet\ \bullet\ \bullet\ a_1: & \bullet\ \bullet\ \bullet\ \bullet\ a_1, & \bullet\ \bullet\ \bullet\ \bullet\ a_1; \\
\bullet\ \bullet\ b_1 & \bullet\ \bullet\ b_1 & \bullet\ \bullet & \bullet\ \bullet\ b_1 & \bullet\ \bullet\ b_1 \\
\bullet\ \bullet\ c_1 & \bullet\ \bullet & \bullet\ \bullet & \bullet\ \bullet\ c_1 & \bullet\ \bullet \\
\bullet\ a_2 & \bullet\ a_2 & \bullet\ a_2 & \bullet & \bullet\ c_1 \\
b_2 & b_2\, c_1 & b_1\, b_2 & a_2 & a_2 \\
c_2 & c_2 & c_1\, c_2 & b_2 & b_2 \\
 & & & c_2 & c_2
\end{array}$$

$$\begin{array}{lllll}
\bullet\ \bullet\ \bullet\ \bullet\ a_1: & \bullet\ \bullet\ \bullet\ \bullet, & \bullet\ \bullet\ \bullet\ \bullet; & \bullet\ \bullet\ \bullet\ \bullet, & \bullet\ \bullet\ \bullet\ \bullet; \\
\bullet\ \bullet & \bullet\ \bullet\ a_1\, a_2 & \bullet\ \bullet\ a_1\, a_2 & \bullet\ \bullet\ a_1\, a_2 & \bullet\ \bullet\ a_1\, a_2 \\
\bullet\ \bullet & \bullet\ \bullet\ b_1\, b_2 & \bullet\ \bullet\ b_1\, b_2 & \bullet\ \bullet\ b_1 & \bullet\ \bullet\ b_1 \\
\bullet\ b_1 & \bullet\ c_1\, c_2 & \bullet\ c_1 & \bullet\ b_2\, c_1 & \bullet\ b_2 \\
a_2\, c_1 & & c_2 & c_2 & c_1\, c_2 \\
b_2 & & & & \\
c_2 & & & &
\end{array}$$

$$\begin{array}{lllll}
\bullet\ \bullet\ \bullet\ \bullet; & \bullet\ \bullet\ \bullet\ \bullet: & \bullet\ \bullet\ \bullet\ \bullet, & \bullet\ \bullet\ \bullet\ \bullet; & \bullet\ \bullet\ \bullet\ \bullet: \\
\bullet\ \bullet\ a_1\, a_2 & \bullet\ \bullet\ a_1\, a_2 & \bullet\ \bullet\ a_1 & \bullet\ \bullet\ a_1 & \bullet\ \bullet\ a_1 \\
\bullet\ \bullet\ b_1 & \bullet\ \bullet & \bullet\ \bullet\ b_1 & \bullet\ \bullet\ b_1 & \bullet\ \bullet \\
\bullet\ c_1 & \bullet\ b_1 & \bullet\ a_2\, c_1 & \bullet\ a_2 & \bullet\ a_2 \\
b_2 & b_2\, c_1 & b_2 & b_2\, c_1 & b_1\, b_2 \\
c_2 & c_2 & c_2 & c_2 & c_1\, c_2
\end{array}$$

$$\begin{array}{llll} \bullet & \bullet & \bullet & \bullet; \\ \bullet & \bullet & a_1 & \\ \bullet & \bullet & b_1 & \\ \bullet & c_1 & & \\ a_2 & & & \\ b_2 & & & \\ c_2 & & & \end{array} \qquad \begin{array}{llll} \bullet & \bullet & \bullet & \bullet: \\ \bullet & \bullet & a_1 & \\ \bullet & \bullet & & \\ \bullet & b_1 & & \\ a_2 & c_1 & & \\ b_2 & & & \\ c_2 & & & \end{array} \qquad \begin{array}{llll} \bullet & \bullet & \bullet & \bullet. \\ \bullet & \bullet & & \\ \bullet & \bullet & & \\ \bullet & a_1 & & \\ a_2 & b_1 & & \\ b_2 & c_1 & & \\ c_2. & & & \end{array}$$

While the number of tableaux which it is necessary to construct in a given product may not be small, nevertheless after a little practice the application of *LR* becomes quite mechanical and is entirely elementary. Each tableau representing an irreducible component, we have the equation

$$\begin{aligned} \{4, 2^2, 1\} \times \{2^3\} = {} & (6, 4^2, 1) + (6, 4, 3, 2) + (6, 4, 3, 1^2) + (6, 4, 2^2, 1) \\ & + (6, 3^2, 2, 1) + (6, 3, 2^3) + (6, 3^2, 1^3) + (6, 3, 2^2, 1^2) \\ & + (6, 2^4, 1) + (5, 4^2, 2) + (5, 4^2, 1^2) + (5, 4, 3^2) \\ & + 2(5, 4, 3, 2, 1) + (5, 4, 2^3) + (5, 4, 3, 1^3) \\ & + (5, 4, 2^2, 1^2) + (5, 3^3, 1) + (5, 3^2, 2^2) + (5, 3^2, 2, 1^2) \\ & + (5, 3, 2^3, 1) + (5, 3^2, 2, 1^2) + (5, 3, 2^3, 1) + (5, 2^5) \\ & + (5, 3^2, 1^4) + (5, 3, 2^2, 1^3) + (5, 2^4, 1^2) + (4^3, 3) \\ & + (4^3, 2, 1) + (4^2, 3^2, 1) + (4^2, 3, 2^2) + (4^2, 3, 2, 1^2) \\ & + (4^2, 2^3, 1) + (4, 3^3, 1^2) + (4, 3^2, 2^2, 1) + (4, 3, 2^4) \\ & + (4, 3^2, 2, 1^3) + (4, 3, 2^3, 1^2) + (4, 2^5, 1). \end{aligned}$$

THE UNIVERSITY OF TORONTO.

REFERENCES.

[1] R. Brauer, "On algebras which are connected with the semi-simple continuous groups," *Annals of Mathematics* (2), vol. 38 (1937), pp. 857-872.

[2] G. Frobenius and I. Schur, "Über die reellen Darstellungen der endlichen Gruppen," *Sitzungsberichte der Preussischen Akademie* 1906, pp. 186-208.

[3] D. E. Littlewood and A. R. Richardson, "Group characters and algebra," *Philosophical Transactions of the Royal Society of London* (A), vol. 233 (1934), pp. 99-141.

[4] P. A. MacMahon, *Combinatory Analysis*, Cambridge, 1915, vols. I and II.

[5] F. D. Murnaghan, "On the representations of the symmetric group," *American Journal of Mathematics*, vol. 59 (1937), pp. 437-488.

[6] ———, "The characters of the symmetric group," *ibid.*, pp. 739-753.

[7] ———, "On the direct product of the irreducible representations of the symmetric group," *Proceedings of the National Academy of Sciences*, vol. 23 (1937), pp. 288-490.

[8] ———, "The analysis of the direct product of irreducible representations of the symmetric group," *American Journal of Mathematics*, vol. 60 (1938), pp. 44-65.

[9] G. de B. Robinson, "Note on an equation of quantitative substitutional analysis," *Proceedings of the London Mathematical Society* (2), vol. 38 (1935), pp. 414-416.

[10] ———, "On the fundamental region of an orthogonal representation of a finite group," *Proceedings of the London Mathematical Society* (2), vol. 43 (1937), pp. 289-301.

[11] I. Schur, *Dissertation*, Berlin, 1901.

[12] ———, "Über die Darstellung der symmetrischen Gruppe durch lineare homogene Substitutionen," *Sitzungsberichte der Preussischen Akademie* 1908, pp. 664-678.

[13] ———, "Über die rationalen Darstellungen der allgemeinen linearen Gruppe," *Sitzungsberichte der Preussischen Akademie* 1927, pp. 58-75.

[14] W. Specht, "Die irreduciblen Darstellungen der symmetrischen Gruppe," *Mathematische Zeitschrift*, vol. 39 (1935), pp. 696-711.

[15] ———, "Darstellungstheorie der Hyperoktaedergruppe," *Mathematische Zeitschrift*, vol. 42 (1937), pp. 629-640.

[16] ———, "Zur Darstellungstheorie der symmetrischen Gruppe," *ibid.*, pp. 774-779.

[17] ———, "Darstellungstheorie der affinen Gruppe," *Mathematische Zeitschrift*, vol. 43 (1937), pp. 120-160.

[18] ———, "Darstellungstheorie der alternierenden Gruppe," *ibid.*, pp. 553-572.

[19] B. L. van der Waerden, "Der Zusammenhang zwischen den Darstellungen der symmetrischen und der linearen Gruppen," *Mathematische Annalen*, vol. 104 (1931), pp. 92-95 also 800.

[20] H. Weyl, *The Theory of Groups and Quantum Mechanics*, New York, 1932.

[21] A. Young, "Quantitative substitutional analysis," Part I, *Proceedings of the London Mathematical Society* (1), vol. 33 (1901), pp. 97-146; Part II, *ibid.* (1), vol. 34 (1902), 361-397; Part III, *ibid.* (2), vol. 28 (1928), pp. 255-292; Part IV, *ibid.* (2), vol. 31 (1930), pp. 253-272; Part V, *ibid.* (2), vol. 31 (1930), pp. 273-288; Part VI, *ibid.* (2), vol. 34 (1932), pp. 196-230; Part VII, *ibid.* (2), vol. 36 (1933), pp. 304-368; Part VIII, *ibid.* (2), vol. 37 (1934), pp. 441-495.

For other references cf. [5].

THE ANALYSIS OF THE KRONECKER PRODUCT OF IRREDUCIBLE REPRESENTATIONS OF THE SYMMETRIC GROUP.*

By F. D. MURNAGHAN.

Introduction. The well known theorems which are known as the Clebsch-Gordan series, and which are fundamental in invariant theory and in the application of group theory to quantum mechanics, are two in number. The first furnishes the analysis of the Kronecker product of any two irreducible representations of the two-dimensional unimodular linear group into its irreducible components; and the second furnishes the analysis of the Kronecker product of any two irreducible representations of the three-dimensional rotation group. The irreducible representations (rational-integral) of the n-dimensional linear group are each described by $j \leq n$, positive integers (whose sum is the degree of the representation in question); hence when $n = 2$ j has only two possible values 1 and 2 (the identity representation, of degree zero, is exceptional, being characterised by the number zero). A further simplification is introduced by the demand that the two-dimensional linear group be *unimodular:* in this case each irreducible representation (continuous) is rational-integral and is characterised by a single label (a positive integer or zero). Thus the first of the two classical Clebsch-Gordan series furnishes the analysis of the Kronecker product $(m) \times (n)$ of two irreducible representations (m) and (n) of the two-dimensional unimodular linear group, each term in the analysis being described by a single non-negative integer. The irreducible representations of the n-dimensional rotation group are each described by $j \leq k$ positive integers where $k = n/2$ if n is even and $k = (n-1)/2$ when n is odd (the identity representation, which is characterised by the number zero, being again exceptional). Hence when $n = 3$ each irreducible representation is again characterised by a single label and the second of the two classical Clebsch-Gordan series furnishes the analysis of the Kronecker product $(m) \times (n)$ of two irreducible representations (m) and (n) of the three-dimensional rotation group, each term in the analysis being described by a single non-negative integer. In a recent paper (1) Brauer describes a method for finding the analysis of the Kronecker product of two irreducible representations of a semi-simple continuous group and makes the following remark: "Dans le cas d'un groupe d'ordre fini, on ne connait pas de loi explicite de

---

* Received May 10, 1938.

cette nature." We give in this paper the analysis of the Kronecker product of various pairs of irreducible representations of the symmetric group on n letters. Each irreducible representation of this group is described by $j \leq n$ positive integers, whose sum is n, so that a restriction analogous to that imposed in the classical Clebsch-Gordan series would be to demand that $n = 2$ which would reduce the question to a triviality. Fortunately no such restriction is necessary: the coefficients in the analysis are *independent of n.* This is the central result of the present paper and we think it well to add a few words of explanation here to make what is meant quite clear. The simplest non-trivial instance of the theory furnishes the analysis of $(n-1,1) \times (n-1,1)$ i. e. the Kronecker square of the irreducible representation $(n-1,1)$ of the symmetric group on n letters (the representation (n), described by the *single* positive integer n, being the identity representation, acts as a unit in the formation of Kronecker products). The result, *valid for arbitrary n,* is

$$(n-1,1) \times (n-1,1) = (n) + (n-1,1) + (n-2,2) + (n-2,1^2).$$

However, one must understand that if a partition of n which occurs in the analysis is disordered, i. e. not in normal non-increasing order, it must be restored to this normal order, or discarded, in the manner described in detail in our previous paper (2) p. 461. Thus when $n = 2$ we obtain on the right the two disordered terms $(0, 2)$ and $(0, 1^2)$; of these the first must be replaced by $-(1^2)$ and the second discarded (because it contains an element, 1, which is greater than the preceding element, 0, by unity). Hence $(1^2) \times (1^2) = (2) + (1^2) - (1^2) = (2)$ which is the trivially evident fact that the Kronecker square of the alternating representation of the symmetric group on two letters is the identity representation of this group. When $n = 3$ we obtain amongst the four terms in the analysis the term $(1, 2)$ which must be discarded so that

$$(2,1) \times (2,1) = (3) + (2,1) + (1^3).$$

For higher values of n no disordered partitions of n occur in the analysis. Thus

$$n = 4;\ (3,1) \times (3,1) = (4) + (3,1) + (2^2) + (2,1^2)$$
$$n = 5;\ (4,1) \times (4,1) = (5) + (4,1) + (3,2) + (3,1^2)$$

and so on. We give in the present paper the analyses of $(n-p, \lambda_2, \cdots) \times (n-q, \mu_2, \cdots)$ for the following values of p and q: $p = 1$, $q = 1, 2, 3, 4, 5$ (there being 18 of these); $p = 2$, $q = 2, 3, 4$ (there being 19 of these); $p = 3$, $q = 3, 4$ (there being 21 of these). It is not surprising that the formulae become increasingly complicated as $p + q$ increases; in fact, the analysis of

$(n-p, \lambda_2, \cdots) \times (n-q, \mu_2, \cdots)$ does not contain any component whose initial element $< n-p-q$. Those we give suffice (in combination with certain elementary and obvious devices) to give the analysis of all products provided $n \leq 8$ and we append these analyses. It would have been possible to construct these analyses from the character tables (by a Fourier analysis method) but such a method is exceedingly tedious and anyone attempting it inevitably feels himself classed amongst the "hewers of wood and drawers of water." By availing ourselves of the two central facts

(a) that the coefficients of the analysis are independent of n and

(b) that the analysis of $(n-p, \cdots) \times (n-q, \cdots)$ does not go deeper than the terms $(n-p-q, \cdots)$

we are able to dispense with the troublesome Fourier analysis. The page references in the following paragraph refer to our paper (**2**). We close this introduction with the doubtless unnecessary remark that the Kronecker products discussed here must not be confused with the "direct" products treated in a previous paper (**3**). The direct product of two irreducible representations of the symmetric groups on m and n letters, respectively, may be regarded as a Kronecker product but *not* of representations of the symmetric groups; we have to take the *attached* irreducible representations, of degrees m and n, respectively, of the full linear group and then the Kronecker product, of degree $m+n$, of these irreducible representations of the full linear group (*not* the symmetric group) is *attached* to the direct product of the two given representations of the symmetric group.

1. **Description of the method.** The essential fact that the coefficients of the analysis of $(n-p, \lambda_2, \cdots) \times (n-q, \mu_2, \cdots)$ depend only on $(\lambda_2, \cdots)$, $(\mu_2, \cdots)$, whose sums are p and q, respectively, and *not* on n is an immediate consequence of a remark made in a previous paper (**4**) in which we gave explicit formulae for the characters of various representations $(n-p, \lambda_2, \cdots)$ of the symmetric group on n letters in terms of the class numbers $(\alpha) = (\alpha_1, \alpha_2, \cdots)$. We pointed out there that the formulae of this type did not involve n explicitly; they are in fact polynomials in $(\alpha_1, \alpha_2, \cdots)$ with coefficients depending on $(\lambda_2, \cdots)$. Since the characters of a Kronecker product are simply the products of the characters of the factors the fact that the coefficients occurring in the analysis of our Kronecker product are independent of n is a foregone conclusion. The rule for rearranging disordered partitions of n is an immediate consequence of the following fact (p. 460): if $(\lambda) = (\lambda_1, \cdots)$ is any partition of n the characteristic $\phi_{(\lambda)}(s)$ of the irreducible representation (λ) is a determinant with the property

$$\phi_{(\ldots\lambda_p,\,\lambda_{p+1},\ldots)}(s) = -\phi_{(\ldots\lambda_{p+1}-1,\,\lambda_p+1\ldots)}(s).$$

Since

$$\phi_{(\lambda)}(s) = \frac{1}{n!}\sum n_{(\alpha)}\chi^{(\alpha)}_{(\lambda)}s^{(\alpha)}$$

it follows that

$$\chi^{(\alpha)}_{(\ldots\lambda_p,\,\lambda_{p+1},\ldots)} = -\chi^{(\alpha)}_{(\ldots\lambda_{p+1}-1,\,\lambda_p+1,\ldots)}$$

(Cf. end of § 1, p. 742, of our previous paper (4)).

It remains to show that the product $(n-p, \lambda_2, \cdots) \times (n-q, \mu_2, \cdots)$ does not contain any terms whose first element is less than $n-p-q$. To do this we observe that the characters of $\Delta(n)$, the identity representation, are 1; that those of $\Delta(n-1, 1)$, whose characteristic is $q_{n-1}(s)s_1$, are α_1 (see (4) p. 740); that those of $\Delta(n-2, 1^2)$, whose characteristic is $q_{n-2}(s)s_1^2$, are $\alpha_1(\alpha_1-1)$ and so on. Since any polynomial in α_1 is a linear combination of the quantities $1, \alpha_1, \alpha_1(\alpha_1-1), \cdots$, it follows that any polynomial in α_1 of degree m is a linear combination of the characters of the representations $\Delta(n-j, 1^j)$, $j = 0, 1, \cdots, m$. Again the representation $\Delta(n-2, 2)$, whose characteristic is $q_{n-2}(s)q_2(s) = \frac{1}{2}q_{n-2}(s)(s_1^2 + s_2)$ has $\frac{1}{2}\alpha_1(\alpha_1-1) + \alpha_2$ for its characters so that α_2 is a linear combination of the characters of representations which go no deeper than $\Delta(n-2, \cdots)$. On considering $\Delta(n-2-j, 2, 1^j)$ whose characteristic is

$$q_{n-2-j}(s)q_2(s)q_1^j(s) = \tfrac{1}{2}q_{n-2-j}(s)(s_1^{j+2} + s_1^j s_2)$$

we see that the product of any polynomial in α_1 of degree m by α_2 is a linear combination of the characters of representations $\Delta(\lambda)$, the representations involved going no deeper than $\Delta(n-m-2, \cdots)$. It would be pedantic to continue the argument; it being now clear that any polynomial in $(\alpha_1, \cdots, \alpha_n)$ whose term of greatest weight is $k \leq n$ (the weight w of any term $\alpha_1^{p_1}\cdots\alpha_n^{p_n}$ being defined by $p_1 + 2p_2 + \cdots + np_n = w$) is a linear combination of the characters of representations $\Delta(\lambda)$, the representations involved going no deeper than $\Delta(n-k, \cdots)$. But the expression for the characters of $D(n-p, \cdots)$ has no term of weight $> p$; hence the expression for the characters of the product $D(n-p, \cdots) \times D(n-q, \cdots)$ has no term of weight $> p+q$. In other words the compound representation $D(n-p, \cdots) \times D(n-q, \cdots)$ is a linear combination (in fact with integral, but not necessarily positive, coefficients) of the (compound) representations $\Delta(\lambda)$, no representations deeper than $\Delta(n-p-q, \cdots)$ being involved. This implies (cf. (2) pp. 469-479) that the product $D(n-p, \cdots) \times D(n-q, \cdots)$ is a linear combination (with positive integral coefficients) of irreducible representations $D(\lambda)$, no representations deeper than $D(n-p-q, \cdots)$ being involved.

We shall illustrate in detail the method of determining the coefficients occurring in the desired analysis by considering the first two instances. We first observe that if we write $(\lambda)\times(\mu) = g^{(\alpha)}_{(\lambda)(\mu)}(\alpha)$ (summation convention!) the coefficient symbol $g^{(\alpha)}_{(\lambda)(\mu)}$ is symmetric in its lower labels:

$$g^{(\alpha)}_{(\lambda)(\mu)} = g^{(\alpha)}_{(\mu)(\lambda)}$$

simply because Kronecker multiplication is commutative. More than this $g^{(\alpha)}_{(\lambda)(\mu)}$ is symmetric in all three labels (λ), (μ), (α). In fact the relation $(\lambda)\times(\mu) = g^{(\alpha)}_{(\lambda)(\mu)}(\alpha)$ which serves to define the integers $g^{(\alpha)}_{(\lambda)(\mu)}$ forces

$$\chi^{(j)}_{(\lambda)}\chi^{(j)}_{(\mu)} = g^{(\alpha)}_{(\lambda)(\mu)}\chi^{(j)}_{(\alpha)}$$

where (j) is any class of the symmetric group on n letters. On multiplication by $n_j\chi^{(j)}_{(\nu)} \div n!$ and summation over all classes (j) we obtain (owing to the orthogonality relations amongst the characters and the fact that they are all real)

$$g^{(\nu)}_{(\lambda)(\mu)} = \frac{1}{n!}\sum_{(j)}\chi^{(j)}_{(\lambda)}\chi^{(j)}_{(\mu)}\chi^{(j)}_{(\nu)}$$

proving the symmetry of the coefficient symbol $g^{(\nu)}_{(\lambda)(\mu)}$ in all three labels (λ), (μ), (ν). In other words the coefficient of (ν) in the analysis of $(\lambda)\times(\mu)$ is the same as the coefficient of (μ) in the analysis of $(\lambda)\times(\nu)$. If in particular (ν) is the identity representation (n) we see that (n) does not appear in any Kronecker product $(\lambda)\times(\mu)$ unless it is a Kronecker square $(\lambda)\times(\lambda)$ and in this case it occurs exactly once (for $(\lambda)\times(n) = (\lambda)$).

These remarks are sufficient to enable us to proceed with our, now trivial, calculations. Thus to analyse $(n-1,1)\times(n-1,1)$ we write it as follows:

$$(n-1,1)\times(n-1,1) = (n) + c_1(n-1,1) + c_2(n-2,2) + c_{1^2}(n-2,1^2)$$

our problem being the determination of the three coefficients c_1, c_2, c_{1^2} which are independent of n. Setting $n = 0$ we find $c_1 = c_{1^2}$, since

$$(-1,1) = -(0),\ (-2,2) = -(1,-1) = 0,\ (-2,1^2) = -(0,-1,1) = (0).$$

There is of course no symmetric group when $n = 0$; but $q_0(s) = 1$ and we may regard (0) as 1. Setting $n = 1$ we obtain $c_2 = 1$; for $(0,1)$ vanishes as also does $(-1,1^2)$ since this last contains an element (the third) which is greater by 2 than an element (the first) two steps ahead of it. Finally setting $n = 2$ we find $c_1 = c_2$; for the product of (1^2) by (1^2) is the associate, namely (2), of (1^2) and $(0,1^2)$ vanishes. Hence $c_1 = c_2 = c_{1^2} = 1$ so that the desired analysis is

$$(n-1,1)\times(n-1,1)=(n)+(n-1,1)+(n-2,2)+(n-2,1^2).$$

In proceeding with the next example $(n-1,1)\times(n-2,2)$ we know that the coefficient of (n) is zero whilst that of $(n-1,1)$ is 1. Hence we write

$$(n-1,1)\times(n-2,2)=(n-1,1)+c_2(n-2,2)+c_{1^2}(n-2,1^2)$$
$$+c_3(n-3,3)+c_{2,1}(n-3,2,1)+c_{1^3}(n-3,1^3).$$

Setting $n=0$ we get $c_{1^2}-c_{1^3}=1$, since $(-2,2)$ vanishes, as do also $(-3,3)$, $(-3,2,1)$, whilst $(-1,1)=-(0)$, $(-2,1^2)=(0)$, $(-3,1^3)=-(0)$. Setting $n=1$ we find $c_2=c_{2,1}$ since $(0,1)$ vanishes, as do also $(-1,1^2)$, $(-2,3)$, $(-2,1^3)$, whilst $(-1,2)=-(1)$ and $(-2,2,1)=(1)$. Setting $n=2$ the left hand side becomes the associate of $(0,2)=-(1^2)$ i. e. $-(2)$. The right hand side becomes $(1-c_2)(1^2)-c_3(2)$ and since the simple characteristics of the symmetric group on two letters are linearly independent we obtain the *two* relations $c_2=1$; $c_3=1$; setting $n=3$ (in which case the factor $(1,2)$ on the left vanishes) we obtain the two relations: $c_3=1$, $c_{2,1}=c_{1^2}$ so that $c_2=1$, $c_{1^2}=1$, $c_3=1$, $c_{2,1}=1$, $c_{1^3}=0$. The various analyses given in the next section were obtained in this way.

2. The results of the analysis. Denoting by $(\lambda_1,\lambda_2,\cdots,\lambda_j)$ the irreducible representation $D(\lambda)$ of the symmetric group on $n=\lambda_1+\lambda_2+\cdots+\lambda_j$ letters we have:

1). $(n-1,1)\times(n-1,1)=(n)+(n-1,1)+(n-2,2)+(n-2,1^2)$.

2). $(n-1,1)\times(n-2,2)=(n-1,1)+(n-2,2)+(n-2,1^2)$
$+(n-3,3)+(n-3,2,1)$.

3). $(n-1,1)\times(n-2,1^2)=(n-1,1)+(n-2,2)+(n-2,1^2)$
$+(n-3,2,1)+(n-3,1^3)$.

4). $(n-1,1)\times(n-3,3)=(n-2,2)+(n-3,3)+(n-3,2,1)$
$+(n-4,4)+(n-4,3,1)$.

5). $(n-1,1)\times(n-3,2,1)=(n-2,2)+(n-2,1^2)+(n-3,3)$
$+2(n-3,2,1)+(n-3,1^3)$
$+(n-4,3,1)+(n-4,2^2)$
$+(n-4,2,1^2)$.

6). $(n-1,1)\times(n-3,1^3)=(n-2,1^2)+(n-3,2,1)+(n-3,1^3)$
$+(n-4,2,1^2)+(n-4,1^4)$.

7). $(n-1,1)\times(n-4,4)=(n-3,3)+(n-4,4)+(n-4,3,1)$
$+(n-5,5)+(n-5,4,1)$.

8). $(n-1,1)\times(n-4,3,1)=(n-3,3)+(n-3,2,1)+(n-4,4)$
$+2(n-4,3,1)+(n-4,2^2)$

$+ (n-4, 2, 1^2) + (n-5, 4, 1)$
$+ (n-5, 3, 2) + (n-5, 3, 1^2)$.

9). $(n-1, 1) \times (n-4, 2^2) = (n-3, 2, 1) + (n-4, 3, 1) + (n-4, 2^2)$
$+ (n-4, 2, 1^2) + (n-5, 3, 2)$
$+ (n-5, 2^2, 1)$.

10). $(n-1, 1) \times (n-4, 2, 1^2) = (n-3, 2, 1) + (n-3, 1^3)$
$+ (n-4, 3, 1) + (n-4, 2^2)$
$+ 2(n-4, 2, 1^2) + (n-4, 1^4)$
$+ (n-5, 3, 1^2) + (n-5, 2^2, 1)$
$+ (n-5, 2, 1^3)$.

11). $(n-1, 1) \times (n-4, 1^4) = (n-3, 1^3) + (n-4, 2, 1^2) + (n-4, 1^4)$
$+ (n-5, 2, 1^3) + (n-5, 1^5)$.

12). $(n-1, 1) \times (n-5, 5) = (n-4, 4) + (n-5, 5) + (n-5, 4, 1)$
$+ (n-6, 6) + (n-6, 5, 1)$

13). $(n-1, 1) \times (n-5, 4, 1) = (n-4, 4) + (n-4, 3, 1) + (n-5, 5)$
$+ 2(n-5, 4, 1) + (n-5, 3, 2)$
$+ (n-5, 3, 1^2) + (n-6, 5, 1)$
$+ (n-6, 4, 2) + (n-6, 4, 1^2)$

14). $(n-1, 1) \times (n-5, 3, 2) = (n-4, 3, 1) + (n-4, 2^2) + (n-5, 4, 1)$
$+ 2(n-5, 3, 2) + (n-5, 3, 1^2)$
$+ (n-5, 2^2, 1) + (n-6, 4, 2)$
$+ (n-6, 3^2) + (n-6, 3, 2, 1)$.

15). $(n-1, 1) \times (n-5, 3, 1^2) = (n-4, 3, 1) + (n-4, 2, 1^2)$
$+ (n-5, 4, 1) + (n-5, 3, 2)$
$+ 2(n-5, 3, 1^2) + (n-5, 2^2, 1)$
$+ (n-5, 2, 1^3) + (n-6, 4, 1^2)$
$+ (n-6, 3, 2, 1) + (n-6, 3, 1^3)$.

16). $(n-1, 1) \times (n-5, 2^2, 1) = (n-4, 2^2) + (n-4, 2, 1^2) + (n-5, 3, 2)$
$+ (n-5, 3, 1^2) + 2(n-5, 2^2, 1)$
$+ (n-5, 2, 1^3) + (n-6, 3, 2, 1)$
$+ (n-6, 2^3) + (n-6, 2^2, 1^2)$.

17). $(n-1, 1) \times (n-5, 2, 1^3) = (n-4, 2, 1^2) + (n-4, 1^4) + (n-5, 3, 1^2)$
$+ (n-5, 2^2, 1) + 2(n-5, 2, 1^3)$
$+ (n-5, 1^5) + (n-6, 3, 1^3)$
$+ (n-6, 2^2, 1^2) + (n-6, 2, 1^4)$.

18). $(n-1, 1) \times (n-5, 1^5) = (n-4, 1^4) + (n-5, 2, 1^3) + (n-5, 1^5)$
$+ (n-6, 2, 1^4) + (n-6, 1^6)$.

19). $$\begin{aligned}(n-2,2)\times(n-2,2) = {} & (n)+(n-1,1)+2(n-2,2)+(n-2,1^2)\\ & +(n-3,3)+2(n-3,2,1)\\ & +(n-3,1^3)+(n-4,4)\\ & +(n-4,3,1)+(n-4,2^2).\end{aligned}$$

20). $$\begin{aligned}(n-2,2)\times(n-2,1^2) = {} & (n-1,1)+(n-2,2)+2(n-2,1^2)\\ & +(n-3,3)+2(n-3,2,1)\\ & +(n-3,1^3)+(n-4,3,1)\\ & +(n-4,2,1^2).\end{aligned}$$

21). $$\begin{aligned}(n-2,2)\times(n-3,3) = {} & (n-1,1)+(n-2,2)+(n-2,1^2)\\ & +2(n-3,3)+2(n-3,2,1)\\ & +(n-4,4)+2(n-4,3,1)\\ & +(n-4,2^2)+(n-4,2,1^2)\\ & +(n-5,5)+(n-5,4,1)\\ & +(n-5,3,2).\end{aligned}$$

22). $$\begin{aligned}(n-2,2)\times(n-3,2,1) = {} & (n-1,1)+2(n-2,2)+2(n-2,1^2)\\ & +2(n-3,3)+4(n-3,2,1)\\ & +2(n-3,1^3)+(n-4,4)\\ & +3(n-4,3,1)+2(n-4,2^2)\\ & +3(n-4,2,1^2)+(n-4,1^4)\\ & +(n-5,4,1)+(n-5,3,2)\\ & +(n-5,3,1^2)+(n-5,2^2,1).\end{aligned}$$

23). $$\begin{aligned}(n-2,2)\times(n-3,1^3) = {} & (n-2,2)+(n-2,1^2)+2(n-3,2,1)\\ & +2(n-3,1^3)+(n-4,3,1)\\ & +(n-4,2^2)+2(n-4,2,1^2)\\ & +(n-4,1^4)+(n-5,3,1^2)\\ & +(n-5,2,1^3).\end{aligned}$$

24). $$\begin{aligned}(n-2,2)\times(n-4,4) = {} & (n-2,2)+(n-3,3)+(n-3,2,1)\\ & +2(n-4,4)+2(n-4,3,1)\\ & +(n-4,2^2)+(n-5,5)\\ & +2(n-5,4,1)+(n-5,3,2)\\ & +(n-5,3,1^2)+(n-6,6)\\ & +(n-6,5,1)+(n-6,4,2).\end{aligned}$$

25). $$\begin{aligned}(n-2,2)\times(n-4,3,1) = {} & (n-2,2)+(n-2,1^2)+2(n-3,3)\\ & +3(n-3,2,1)+(n-3,1^3)\\ & +2(n-4,4)+5(n-4,3,1)\\ & +2(n-4,2^2)+3(n-4,2,1^2)\\ & +(n-5,5)+3(n-5,4,1)\\ & +3(n-5,3,2)+3(n-5,3,1^2)\\ & +2(n-5,2^2,1)+(n-5,2,1^3)\end{aligned}$$

$$\begin{aligned}
&+(n-6,5,1)+(n-6,4,2)\\
&+(n-6,4,1^2)+(n-6,3^2)\\
&+(n-6,3,2,1).
\end{aligned}$$

26). $(n-2,2)\times(n-4,2^2)=(n-2,2)+(n-3,3)+2(n-3,2,1)$

$$\begin{aligned}
&+(n-3,1^3)+(n-4,4)\\
&+2(n-4,3,1)+3(n-4,2^2)\\
&+2(n-4,2,1^2)+(n-4,1^4)\\
&+(n-5,4,1)+2(n-5,3,2)\\
&+2(n-5,3,1^2)+2(n-5,2^2,1)\\
&+(n-5,2,1^3)+(n-6,4,2)\\
&+(n-6,3,2,1)+(n-6,2^3).
\end{aligned}$$

27). $(n-2,2)\times(n-4,2,1^2)=(n-2,1^2)+(n-3,3)+3(n-3,2,1)$

$$\begin{aligned}
&+2(n-3,1^3)+3(n-4,3,1)\\
&+2(n-4,2^2)+5(n-4,2,1^2)\\
&+2(n-4,1^4)+(n-5,4,1)\\
&+2(n-5,3,2)+3(n-5,3,1^2)\\
&+3(n-5,2^2,1)+3(n-5,2,1^3)\\
&+(n-5,1^5)+(n-6,4,1^2)\\
&+(n-6,3,2,1)+(n-6,3,1^3)\\
&+(n-6,2^2,1^2).
\end{aligned}$$

28). $(n-2,2)\times(n-4,1^4)=(n-3,2,1)+(n-3,1^3)+(n-4,2^2)$

$$\begin{aligned}
&+2(n-4,2,1^2)+2(n-4,1^4)\\
&+(n-5,3,1^2)+(n-5,2^2,1)\\
&+2(n-5,2,1^3)+(n-5,1^5)\\
&+(n-6,3,1^3)+(n-6,2,1^4).
\end{aligned}$$

29). $(n-2,1^2)\times(n-2,1^2)=(n)+(n-1,1)+2(n-2,2)$

$$\begin{aligned}
&+(n-2,1^2)+(n-3,3)\\
&+2(n-3,2,1)+(n-3,1^3)\\
&+(n-4,2^2)+(n-4,2,1^2)\\
&+(n-4,1^4).
\end{aligned}$$

30). $(n-2,1^2)\times(n-3,3)=(n-2,2)+(n-2,1^2)+(n-3,3)$

$$\begin{aligned}
&+2(n-3,2,1)+(n-3,1^3)\\
&+(n-4,4)+2(n-4,3,1)\\
&+(n-4,2^2)+(n-4,2,1^2)\\
&+(n-5,4,1)+(n-5,3,1^2).
\end{aligned}$$

31). $(n-2,1^2)\times(n-3,2,1)=(n-1,1)+2(n-2,2)+2(n-2,1^2)$

$$\begin{aligned}
&+2(n-3,3)+4(n-3,2,1)\\
&+2(n-3,1^3)+(n-4,4)\\
&+3(n-4,3,1)+2(n-4,2^2)
\end{aligned}$$

$$\begin{aligned}
&+3(n-4,2,1^2)+(n-4,1^4)\\
&+(n-5,3,2)+(n-5,3,1^2)\\
&+(n-5,2^2,1)+(n-5,2,1^3).
\end{aligned}$$

32).
$$\begin{aligned}
(n-2,1^2)\times(n-3,1^3)=\;&(n-1,1)+(n-2,2)+(n-2,1^2)\\
&+(n-3,3)+2(n-3,2,1)\\
&+(n-3,1^3)+(n-4,3,1)\\
&+(n-4,2^2)+2(n-4,2,1^2)\\
&+(n-4,1^4)+(n-5,2^2,1)\\
&+(n-5,2,1^3)+(n-5,1^5).
\end{aligned}$$

33).
$$\begin{aligned}
(n-2,1^2)\times(n-4,4)=\;&(n-3,3)+(n-3,2,1)+(n-4,4)\\
&+2(n-4,3,1)+(n-4,2,1^2)\\
&+(n-5,5)+2(n-5,4,1)\\
&+(n-5,3,2)+(n-5,3,1^2)\\
&+(n-6,5,1)+(n-6,4,1^2).
\end{aligned}$$

34).
$$\begin{aligned}
(n-2,1^2)\times(n-4,3,1)=\;&(n-2,2)+2(n-3,3)+3(n-3,2,1)\\
&+(n-3,1^3)+2(n-4,4)\\
&+4(n-4,3,1)+3(n-4,2^2)\\
&+3(n-4,2,1^2)+(n-4,1^4)\\
&+(n-5,5)+3(n-5,4,1)\\
&+3(n-5,3,2)+3(n-5,3,1^2)\\
&+2(n-5,2^2,1)+(n-5,2,1^3)\\
&+(n-6,4,2)+(n-6,4,1^2)\\
&+(n-6,3,2,1)+(n-6,3,1^3).
\end{aligned}$$

35).
$$\begin{aligned}
(n-2,1^2)\times(n-4,2^2)=\;&(n-2,1^2)+(n-3,3)+2(n-3,2,1)\\
&+(n-3,1^3)+3(n-4,3,1)\\
&+(n-4,2^2)+3(n-4,2,1^2)\\
&+(n-5,4,1)+2(n-5,3,2)\\
&+2(n-5,3,1^2)+2(n-5,2^2,1)\\
&+(n-5,2,1^3)+(n-6,3^2)\\
&+(n-6,3,2,1)+(n-6,2^2,1^2).
\end{aligned}$$

36).
$$\begin{aligned}
(n-2,1^2)\times(n-4,2,1^2)=\;&(n-2,2)+(n-2,1^2)+(n-3,3)\\
&+3(n-3,2,1)+2(n-3,1^3)\\
&+(n-4,4)+3(n-4,3,1)\\
&+3(n-4,2^2)+4(n-4,2,1^2)\\
&+2(n-4,1^4)+(n-5,4,1)\\
&+2(n-5,3,2)+3(n-5,3,1^2)\\
&+3(n-5,2^2,1)+3(n-5,2,1^3)\\
&+(n-5,1^5)+(n-6,3,2,1)\\
&+(n-6,3,1^3)+(n-6,2^3)\\
&+(n-6,2^2,1^2)+(n-6,2,1^4).
\end{aligned}$$

$$\begin{aligned}
37).\quad (n-2,1^2)\times(n-4,1^4) = {} & (n-2,1^2)+(n-3,2,1)+(n-3,1^3)\\
& +(n-4,3,1)+2(n-4,2,1^2)\\
& +(n-4,1^4)+(n-5,3,1^2)\\
& +(n-5,2^2,1)+2(n-5,2,1^3)\\
& +(n-5,1^5)+(n-6,2^2,1^2)\\
& +(n-6,2,1^4)+(n-6,1^6).
\end{aligned}$$

$$\begin{aligned}
38).\quad (n-3,3)\times(n-3,3) = {} & (n)+(n-1,1)+2(n-2,2)+(n-2,1^2)\\
& +2(n-3,3)+2(n-3,2,1)\\
& +(n-3,1^3)+2(n-4,4)\\
& +3(n-4,3,1)+2(n-4,2^2)\\
& +(n-4,2,1^2)+(n-5,5)\\
& +2(n-5,4,1)+2(n-5,3,2)\\
& +(n-5,3,1^2)+(n-5,2^2,1)\\
& +(n-6,6)+(n-6,5,1)\\
& +(n-6,4,2)+(n-6,3^2).
\end{aligned}$$

$$\begin{aligned}
39).\quad (n-3,3)\times(n-3,2,1) = {} & (n-1,1)+2(n-2,2)+2(n-2,1^2)\\
& +2(n-3,3)+5(n-3,2,1)\\
& +2(n-3,1^3)+2(n-4,4)\\
& +5(n-4,3,1)+3(n-4,2^2)\\
& +4(n-4,2,1^2)+(n-4,1^4)\\
& +(n-5,5)+3(n-5,4,1)\\
& +3(n-5,3,2)+3(n-5,3,1^2)\\
& +2(n-5,2^2,1)+(n-5,2,1^3)\\
& +(n-6,5,1)+(n-6,4,2)\\
& +(n-6,4,1^2)+(n-6,3,2,1).
\end{aligned}$$

$$\begin{aligned}
40).\quad (n-3,3)\times(n-3,1^3) = {} & (n-2,1^2)+(n-3,3)+2(n-3,2,1)\\
& +2(n-3,1^3)+2(n-4,3,1)\\
& +(n-4,2^2)+3(n-4,2,1^2)\\
& +(n-4,1^4)+(n-5,4,1)\\
& +(n-5,3,2)+2(n-5,3,1^2)\\
& +(n-5,2^2,1)+(n-5,2,1^3)\\
& +(n-6,4,1^2)+(n-6,3,1^3).
\end{aligned}$$

$$\begin{aligned}
41).\quad (n-3,3)\times(n-4,4) = {} & (n-1,1)+(n-2,2)+(n-2,1^2)\\
& +2(n-3,3)+2(n-3,2,1)\\
& +2(n-4,4)+3(n-4,3,1)\\
& +(n-4,2^2)+(n-4,2,1^2)\\
& +2(n-5,5)+3(n-5,4,1)\\
& +3(n-5,3,2)+(n-5,3,1^2)
\end{aligned}$$

$$\begin{aligned}
&+(n-5,2^2,1)+(n-6,6)\\
&+2(n-6,5,1)+2(n-6,4,2)\\
&+(n-6,4,1^2)+(n-6,3^2)\\
&+(n-6,3,2,1)+(n-7,7)\\
&+(n-7,6,1)+(n-7,5,2)\\
&+(n-7,4,3).
\end{aligned}$$

$$\begin{aligned}
42).\quad (n-3,3)\times(n-4,3,1)=&\,(n-1,1)+2(n-2,2)+2(n-2,1^2)\\
&+3(n-3,3)+5(n-3,2,1)\\
&+2(n-3,1^3)+3(n-4,4)\\
&+7(n-4,3,1)+4(n-4,2^2)\\
&+5(n-4,2,1^2)+(n-4,1^4)\\
&+2(n-5,5)+6(n-5,4,1)\\
&+6(n-5,3,2)+6(n-5,3,1^2)\\
&+4(n-5,2^2,1)+2(n-5,2,1^3)\\
&+(n-6,6)+3(n-6,5,1)\\
&+4(n-6,4,2)+3(n-6,4,1^2)\\
&+2(n-6,3^2)+4(n-6,3,2,1)\\
&+(n-6,3,1^3)+(n-6,2^3)\\
&+(n-6,2^2,1^2)+(n-7,6,1)\\
&+(n-7,5,2)+(n-7,5,1^2)\\
&+(n-7,4,3)+(n-7,4,2,1)\\
&+(n-7,3^2,1).
\end{aligned}$$

$$\begin{aligned}
43).\quad (n-3,3)\times(n-4,2^2)=&\,(n-2,2)+(n-2,1^2)+2(n-3,3)\\
&+3(n-3,2,1)+(n-3,1^3)\\
&+(n-4,4)+4(n-4,3,1)\\
&+3(n-4,2^2)+4(n-4,2,1^2)\\
&+(n-4,1^4)+(n-5,5)\\
&+3(n-5,4,1)+4(n-5,3,2)\\
&+3(n-5,3,1^2)+4(n-5,2^2,1)\\
&+2(n-5,2,1^3)+(n-5,1^5)\\
&+(n-6,5,1)+2(n-6,4,2)\\
&+2(n-6,4,1^2)+(n-6,3^2)\\
&+3(n-6,3,2,1)+(n-6,3,1^3)\\
&+(n-6,2^3)+(n-6,2^2,1^2)\\
&+(n-7,5,2)+(n-7,4,2,1)\\
&+(n-7,3,2^2).
\end{aligned}$$

$$\begin{aligned}
44).\quad (n-3,3)\times(n-4,2,1^2)=&\,(n-2,2)+(n-2,1^2)+(n-3,3)\\
&+4(n-3,2,1)+3(n-3,1^3)\\
&+(n-4,4)+5(n-4,3,1)\\
&+4(n-4,2^2)+7(n-4,2,1^2)
\end{aligned}$$

$$\begin{aligned}
&+3(n-4,1^4)+3(n-5,4,1)\\
&+4(n-5,3,2)+7(n-5,3,1^2)\\
&+5(n-5,2^2,1)+5(n-5,2,1^3)\\
&+(n-5,1^5)+(n-6,5,1)\\
&+2(n-6,4,2)+3(n-6,4,1^2)\\
&+(n-6,3^2)+4(n-6,3,2,1)\\
&+3(n-6,3,1^3)+(n-6,2^3)\\
&+2(n-6,2^2,1^2)+(n-6,2,1^4)\\
&+(n-7,5,1^2)+(n-7,4,2,1)\\
&+(n-7,4,1^3)+(n-7,3,2,1^2).
\end{aligned}$$

45). $(n-3,3)\times(n-4,1^4)=(n-3,2,1)+(n-3,1^3)+(n-4,3,1)$

$$\begin{aligned}
&+(n-4,2^2)+3(n-4,2,1^2)\\
&+2(n-4,1^4)+(n-5,3,2)\\
&+2(n-5,3,1^2)+2(n-5,2^2,1)\\
&+3(n-5,2,1^3)+(n-5,1^5)\\
&+(n-6,4,1^2)+(n-6,3,2,1)\\
&+2(n-6,3,1^3)+(n-6,2^2,1^2)\\
&+(n-6,2,1^4)+(n-7,4,1^3)\\
&+(n-7,3,1^4).
\end{aligned}$$

46). $(n-3,2,1)\times(n-3,2,1)=(n)+2(n-1,1)+4(n-2,2)$

$$\begin{aligned}
&+4(n-2,1^2)+5(n-3,3)\\
&+9(n-3,2,1)+5(n-3,1^3)\\
&+3(n-4,4)+9(n-4,3,1)\\
&+5(n-4,2^2)+9(n-4,2,1^2)\\
&+3(n-4,1^4)+(n-5,5)\\
&+4(n-5,4,1)+5(n-5,3,2)\\
&+5(n-5,3,1^2)+5(n-5,2^2,1)\\
&+4(n-5,2,1^3)+(n-5,1^5)\\
&+(n-6,4,2)+(n-6,4,1^2)\\
&+(n-6,3^2)+2(n-6,3,2,1)\\
&+(n-6,3,1^3)+(n-6,2^3)\\
&+(n-6,2^2,1^2).
\end{aligned}$$

47). $(n-3,2,1)\times(n-3,1^3)=(n-1,1)+2(n-2,2)+2(n-2,1^2)$

$$\begin{aligned}
&+3(n-3,3)+5(n-3,2,1)\\
&+3(n-3,1^3)+(n-4,4)\\
&+4(n-4,3,1)+3(n-4,2^2)\\
&+5(n-4,2,1^2)+2(n-4,1^4)\\
&+(n-5,4,1)+2(n-5,3,2)\\
&+3(n-5,3,1^2)+3(n-5,2^2,1)\\
&+3(n-5,2,1^3)+(n-5,1^5)
\end{aligned}$$

$$\begin{aligned}
&+(n-6,3,2,1)+(n-6,3,1^3)\\
&+(n-6,2^2,1^2)+(n-6,2,1^4).
\end{aligned}$$

48). $$\begin{aligned}
(n-3,2,1)\times(n-4,4)={}&(n-2,2)+(n-2,1^2)+2(n-3,3)\\
&+3(n-3,2,1)+(n-3,1^3)\\
&+2(n-4,4)+5(n-4,3,1)\\
&+3(n-4,2^2)+3(n-4,2,1^2)\\
&+2(n-5,5)+5(n-5,4,1)\\
&+4(n-5,3,2)+4(n-5,3,1^2)\\
&+2(n-5,2^2,1)+(n-5,2,1^3)\\
&+(n-6,6)+3(n-6,5,1)\\
&+3(n-6,4,2)+3(n-6,4,1^2)\\
&+(n-6,3^2)+2(n-6,3,2,1)\\
&+(n-6,3,1^3)+(n-7,6,1)\\
&+(n-7,5,2)+(n-7,5,1^2)\\
&+(n-7,4,2,1).
\end{aligned}$$

49). $$\begin{aligned}
(n-3,2,1)\times(n-4,3,1)={}&(n-1,1)+3(n-2,2)+3(n-2,1^2)\\
&+5(n-3,3)+9(n-3,2,1)\\
&+4(n-3,1^3)+5(n-4,4)\\
&+13(n-4,3,1)+8(n-4,2^2)\\
&+11(n-4,2,1^2)+3(n-4,1^4)\\
&+3(n-5,5)+10(n-5,4,1)\\
&+11(n-5,3,2)+12(n-5,3,1^2)\\
&+9(n-5,2^2,1)+6(n-5,2,1^3)\\
&+(n-5,1^5)+(n-6,6)\\
&+4(n-6,5,1)+6(n-6,4,2)\\
&+6(n-6,4,1^2)+3(n-6,3^2)\\
&+8(n-6,3,2,1)+4(n-6,3,1^3)\\
&+2(n-6,2^3)+3(n-6,2^2,1^2)\\
&+(n-6,2,1^4)+(n-7,5,2)\\
&+(n-7,5,1^2)+(n-7,4,3)\\
&+2(n-7,4,2,1)+(n-7,4,1^3)\\
&+(n-7,3^2,1)+(n-7,3,2^2)\\
&+(n-7,3,2,1^2).
\end{aligned}$$

50). $$\begin{aligned}
(n-3,2,1)\times(n-4,2^2)+{}&(n-1,\,)+2(n-2,2)+2(n-2,1^2)\\
&+3(n-3,3)+6(n-3,2,1)\\
&+3(n-3,1^3)+3(n-4,4)\\
&+8(n-4,3,1)+5(n-4,2^2)\\
&+8(n-4,2,1^2)+3(n-4,1^4)\\
&+(n-5,5)+5(n-5,4,1)\\
&+7(n-5,3,2)+8(n-5,3,1^2)
\end{aligned}$$

$+ 7(n-5, 2^2, 1) + 5(n-5, 2, 1^3)$
$+ (n-5, 1^5) + (n-6, 5, 1)$
$+ 3(n-6, 4, 2) + 3(n-6, 4, 1^2)$
$+ 2(n-6, 3^2) + 6(n-6, 3, 2, 1)$
$+ 3(n-6, 3, 1^3) + 2(n-6, 2^3)$
$+ 3(n-6, 2^2, 1^2) + (n-6, 2, 1^4)$
$+ (n-7, 4, 3) + (n-7, 4, 2, 1)$
$+ (n-7, 3^2, 1) + (n-7, 3, 2^2)$
$+ (n-7, 3, 2, 1^2) + (n-7, 2^3, 1).$

51). $(n-3, 2, 1) \times (n-4, 2, 1^2) = (n-1, 1) + 3(n-2, 2) + 3(n-2, 1^2)$
$+ 4(n-3, 3) + 9(n-3, 2, 1)$
$+ 5(n-3, 1^3) + 3(n-4, 4)$
$+ 11(n-4, 3, 1) + 8(n-4, 2^2)$
$+ 13(n-4, 2, 1^2) + 5(n-4, 1^4)$
$+ (n-5, 5) + 6(n-5, 4, 1)$
$+ 9(n-5, 3, 2) + 12(n-5, 3, 1^2)$
$+ 11(n-5, 2^2, 1) + 10(n-5, 2, 1^3)$
$+ 3(n-5, 1^5) + (n-6, 5, 1)$
$+ 3(n-6, 4, 2) + 4(n-6, 4, 1^2)$
$+ 2(n-6, 3^2) + 8(n-6, 3, 2, 1)$
$+ 6(n-6, 3, 1^3) + 3(n-6, 2^3)$
$+ 6(n-6, 2^2, 1^2) + 4(n-6, 2, 1^4)$
$+ (n-6, 1^6) + (n-7, 4, 2, 1)$
$+ (n-7, 4, 1^3) + (n-7, 3^2, 1)$
$+ (n-7, 3, 2^2) + 2(n-7, 3, 2, 1^2)$
$+ (n-7, 3, 1^4) + (n-7, 2^3, 1)$
$+ (n-7, 2^2, 1^3).$

52). $(n-3, 2, 1) \times (n-4, 1^4) = (n-2, 2) + (n-2, 1^2) + (n-3, 3)$
$+ 3(n-3, 2, 1) + 2(n-3, 1^3)$
$+ 3(n-4, 3, 1) + 3(n-4, 2^2)$
$+ 5(n-4, 2, 1^2) + 2(n-4, 1^4)$
$+ (n-5, 4, 1) + 2(n-5, 3, 2)$
$+ 4(n-5, 3, 1^2) + 4(n-5, 2^2, 1)$
$+ 5(n-5, 2, 1^3) + 2(n-5, 1^5)$
$+ (n-6, 4, 1^2) + 2(n-6, 3, 2, 1)$
$+ 3(n-6, 3, 1^3) + (n-6, 2^3)$
$+ 3(n-6, 2^2, 1^2) + 3(n-6, 2, 1^4)$
$+ (n-6, 1^6) + (n-7, 3, 2, 1^2)$
$+ (n-7, 3, 1^4) + (n-7, 2^2, 1^3)$
$+ (n-7, 2, 1^5).$

$$\begin{aligned}
53).\ (n-3,1^3)\times(n-3,1^3) = {} & (n)+(n-1,1)+2(n-2,2)\\
& +(n-2,1^2)+2(n-3,3)\\
& +2(n-3,2,1)+(n-3,1^3)\\
& +(n-4,4)+2(n-4,3,1)\\
& +2(n-4,2^2)+2(n-4,2,1^2)\\
& +(n-4,1^4)+(n-5,3,2)\\
& +(n-5,3,1^2)+2(n-5,2^2,1)\\
& +2(n-5,2,1^3)+(n-5,1^5)\\
& +(n-6,2^3)+(n-6,2^2,1^2)\\
& +(n-6,2,1^4)+(n-6,1^6).
\end{aligned}$$

$$\begin{aligned}
54).\ (n-3,1^3)\times(n-4,4) = {} & (n-3,2,1)+(n-3,1^3)+(n-4,4)\\
& +2(n-4,3,1)+(n-4,2^2)\\
& +2(n-4,2,1^2)+(n-4,1^4)\\
& +2(n-5,4,1)+(n-5,3,2)\\
& +3(n-5,3,1^2)+(n-5,2^2,1)\\
& +(n-5,2,1^3)+(n-6,5,1)\\
& +(n-6,4,2)+2(n-6,4,1^2)\\
& +(n-6,3,2,1)+(n-6,3,1^3)\\
& +(n-7,5,1^2)+(n-7,4,1^3).
\end{aligned}$$

$$\begin{aligned}
55).\ (n-3,1^3)\times(n-4,3,1) = {} & (n-2,2)+(n-2,1^2)+2(n-3,3)\\
& +4(n-3,2,1)+2(n-3,1^3)\\
& +2(n-4,4)+6(n-4,3,1)\\
& +4(n-4,2^2)+6(n-4,2,1^2)\\
& +2(n-4,1^4)+(n-5,5)\\
& +4(n-5,4,1)+5(n-5,3,2)\\
& +6(n-5,3,1^2)+5(n-5,2^2,1)\\
& +4(n-5,2,1^3)+(n-5,1^5)\\
& +(n-6,5,1)+2(n-6,4,2)\\
& +3(n-6,4,1^2)+(n-6,3^2)\\
& +4(n-6,3,2,1)+3(n-6,3,1^3)\\
& +(n-6,2^3)+2(n-6,2^2,1^2)\\
& +(n-6,2,1^4)+(n-7,4,2,1)\\
& +(n-7,4,1^3)+(n-7,3,2,1^2)\\
& +(n-7,3,1^4).
\end{aligned}$$

$$\begin{aligned}
56).\ (n-3,1^3)\times(n-4,2^2) = {} & (n-2,2)+(n-2,1^2)+(n-3,3)\\
& +3(n-3,2,1)+2(n-3,1^3)\\
& +(n-4,4)+4(n-4,3,1)\\
& +3(n-4,2^2)+4(n-4,2,1^2)\\
& +(n-4,1^4)+2(n-5,4,1)\\
& +3(n-5,3,2)+5(n-5,3,1^2)
\end{aligned}$$

$$\begin{aligned}
&+3(n-5,2^2,1)+3(n-5,2,1^3)\\
&+(n-6,4,2)+(n-6,4,1^2)\\
&+(n-6,3^2)+3(n-6,3,2,1)\\
&+2(n-6,3,1^3)+(n-6,2^3)\\
&+2(n-6,2^2,1^2)+(n-6,2,1^4)\\
&+(n-7,3^2,1)+(n-7,3,2,1^2)\\
&+(n-7,2^2,1^3).
\end{aligned}$$

57). $(n-3,1^3)\times(n-4,2,1^2)=(n-1,1)+2(n-2,2)+2(n-2,1^2)$

$$\begin{aligned}
&+3(n-3,3)+5(n-3,2,1)\\
&+2(n-3,1^3)+2(n-4,4)\\
&+6(n-4,3,1)+4(n-4,2^2)\\
&+6(n-4,2,1^2)+2(n-4,1^4)\\
&+(n-5,5)+3(n-5,4,1)\\
&+5(n-5,3,2)+5(n-5,3,1^2)\\
&+6(n-5,2^2,1)+5(n-5,2,1^3)\\
&+2(n-5,1^5)+(n-6,4,2)\\
&+(n-6,4,1^2)+(n-6,3^2)\\
&+4(n-6,3,2,1)+3(n-6,3,1^3)\\
&+2(n-6,2^3)+4(n-6,2^2,1^2)\\
&+3(n-6,2,1^4)+(n-6,1^6)\\
&+(n-7,3,2^2)+(n-7,3,2,1^2)\\
&+(n-7,3,1^4)+(n-7,2^3,1)\\
&+(n-7,2^2,1^3)+(n-7,2,1^5).
\end{aligned}$$

58). $(n-3,1^3)\times(n-4,1^4)=(n-1,1)+(n-2,2)+(n-2,1^2)$

$$\begin{aligned}
&+(n-3,3)+2(n-3,2,1)\\
&+(n-3,1^3)+(n-4,4)\\
&+2(n-4,3,1)+(n-4,2^2)\\
&+2(n-4,2,1^2)+(n-4,1^4)\\
&+(n-5,4,1)+(n-5,3,2)\\
&+2(n-5,3,1^2)+2(n-5,2^2,1)\\
&+2(n-5,2,1^3)+(n-5,1^5)\\
&+(n-6,3,2,1)+(n-6,3,1^3)\\
&+(n-6,2^3)+2(n-6,2^2,1^2)\\
&+2(n-6,2,1^4)+(n-6,1^6)\\
&+(n-7,2^3,1)+(n-7,2^2,1^3)\\
&+(n-7,2,1^5)+(n-7,1^7).
\end{aligned}$$

---

3. Tables furnishing the analysis of the Kronecker products for $n \leqq 8$. Since the product of any irreducible representation (λ) of the symmetric group on n letters by the alternating representation (1^n) is the associated irreducible representation $(\lambda)^*$ the product $(\lambda) \times (\mu)$ is the associate of the product $(\lambda) \times (\mu)^*$ or, equivalently, the associate of the product $(\lambda)^* \times (\mu)$. In fact these statements merely reflect the fact that Kronecker multiplication is associative:

$$(\lambda) \times \{(\mu) \times (\nu)\} = \{(\lambda) \times (\mu)\} \times (\nu).$$

For when $(\nu) = (1^n)$ the left-hand side of this equality is $(\lambda) \times (\mu)^*$ whilst the right-hand side is $\{(\lambda) \times (\mu)\}^*$. As a corollary we have $(\lambda)^* \times (\mu)^* = (\lambda) \times (\mu)$; for the left-hand side $= \{(\lambda)^* \times (\mu)\}^* = \{(\lambda) \times (\mu)\}^{**} = (\lambda) \times (\mu)$, the operation of taking the associate of a representation being reflexive. These facts shorten considerably the following tables. The coefficients appear in the body of the tables; if we enter the tables on the left with a product $(\lambda) \times (\mu)$ the representations which are multiplied by the coefficients are found at the *head* of the table. The product $(\lambda) \times (\mu)^*$ is found at the same place in the table but at the right rather than the left and the representations which are multiplied by the coefficients occurring in the table are found at the foot of the table. If a product $(\lambda) \times (\mu)$ is not found in the table either $(\lambda)^* \times (\mu)^*$, which is the same as $(\lambda) \times (\mu)$, or $(\lambda)^* \times (\mu)$, which is the associate of $(\lambda) \times (\mu)$, will be found. The general formulae already given furnish all products up to $n \leqq 6$ and all products for $n = 7$ save $(3^2, 1) \times (3^2, 1)$ and its associate $(3^2, 1) \times (3, 2^2)$. From the symmetry properties of the coefficient symbol $g_{(\lambda)(\mu)}^{(\nu)}$ all coefficients in the analysis of $(3^2, 1) \times (3^2, 1)$ are known, from the other products in the table for $n = 7$, save $c_{3^2,1}$ and $c_{3,2^2}$. These are determined from the character table for $n = 7$ the sum being determined by an even class and the difference by an odd class. For $n = 8$ the formulae given determine all products save those from $(4^2) \times (4^2)$ to the end of the table. The following detail of the analysis of this product will show clearly how all analyses of products which are not furnished by the formulae of section **2** may be obtained. From the symmetry properties of the coefficient symbol we have

$$\begin{aligned}(4^2) \times (4^2) = (8) &+ (6, 2) + (5, 1^3) + c_{4^2}(4^2) + c_{4,3,1}(4, 3, 1) + c_{4,2^2}(4, 2^2) \\ &+ c_{4,2,1^2}(4, 2, 1^2) + c_{3^2,2}(3^2, 2) + c_{3^2,1^2}(3^2, 1^2) \\ &+ c_{3,2^2,1}(3, 2^2, 1) + c_{2^4}(2^4).\end{aligned}$$

From the table of characters (**5**, p. 179) for the symmetric group on 8 letters we find

(a) from the class $(1, 7)$, $c_{4,2,1^2} = 0$.
(b) from the class $(2, 6)$, $\alpha - \beta = 2$ where $\alpha = c_{4,2^2} + c_{3^2,1^2}$, $\beta = c_{4,3,1} + c_{3,2^2,1}$.
(c) from the class $(3, 5)$ $\alpha - \gamma = c_{3^2,2}$ where $\gamma = c_{4^2} + c_{2^4}$.
(d) from the class (2^4), $3\gamma - \beta + 4\alpha - 6c_{3^2,2} = 14$.
(e) from the class (1^8), $\gamma + 5\beta + 4\alpha + c_{3^2,2} = 10$.
Hence $\alpha = 2$, $\beta = 0$, $\gamma = 2$, $c_{3^2,2} = 0$.

(f) from the class $(1^6, 2)$ we find $7\gamma' + 5\beta' + 2\alpha' = 0$ where
$\alpha' = c_{4,2^2} - c_{3^2,1^2}$; $\beta' = c_{4,3,1} - c_{3,2^2,1}$; $\gamma' = c_{4^2} - c_{2^4}$.
(g) from the class $(1^4, 4)$ we find $\gamma' + 2\beta' = 0$.
(h) from the class $(2, 3^2)$ we find $\alpha' + \beta' - 2\gamma' = 0$.
Hence $\alpha' = 0 = \beta' = \gamma'$ so that $c_{4,2^2} = c_{3^2,2} = 1$; $c_{4,3,1} = c_{3,2^2,1} = 0$; $c_{4^2} = c_{2^4} = 1$ and we have the analysis given in the table.

In the following tables the products by the identity representation (n), which acts as a unit, and by the alternating representation (1^n), which sends each representation into its associate, are omitted. Thus the table for $n = 2$ is not given.

1. $n = 3$.

$$(2,1) \times (2,1) = (3) + (2,1) + (1^3).$$

2. $n = 4$.

↑	(4)	$(3,1)$	(2^2)	$(2,1^2)$	(1^4)	↓
$(3,1) \times (3,1)$	1	1	1	1		$(3,1) \times (2,1^2)$
$(3,1) \times (2^2)$		1		1		$(3,1) \times (2^2)$
$(2^2) \times (2^2)$	1		1		1	$(2^2) \times (2^2)$
	(1^4)	$(2,1^2)$	(2^2)	$(3,1)$	(4)	

3. $n = 5$.

↑	(5)	$(4,1)$	$(3,2)$	$(3,1^2)$	$(2^2,1)$	$(2,1^3)$	(1^5)	↓
$(4,1) \times (4,1)$	1	1	1	1				$(4,1) \times (2,1^3)$
$(4,1) \times (3,2)$		1	1	1	1			$(4,1) \times (2^2,1)$
$(4,1) \times (3,1^2)$		1	1	1	1	1		$(4,1) \times (3,1^2)$
$(3,2) \times (3,2)$	1	1	1	1	1	1		$(3,2) \times (2^2,1)$
$(3,2) \times (3,1^2)$		1	1	2	1	1		$(3,2) \times (3,1^2)$
$(3,1^2) \times (3,1^2)$	1	1	2	1	2	1	1	$(3,1^2) \times (3,1^2)$
	(1^5)	$(2,1^3)$	$(2^2,1)$	$(3,1^2)$	$(3,2)$	$(4,1)$	(5)	

4. $n = 6$.

↑	(6)	$(5,1)$	$(4,2)$	$(4,1^2)$	(3^2)	$(3,2,1)$	$(3,1^3)$	(2^3)	$(2^2,1^2)$	$(2,1^4)$	(1^6)	↓
$(5,1) \times (5,1)$	1	1	1	1								$(5,1) \times (2,1^4)$
$(5,1) \times (4,2)$		1	1	1	1	1						$(5,1) \times (2^2,1^2)$
$(5,1) \times (4,1^2)$		1	1	1		1	1					$(5,1) \times (3,1^3)$
$(5,1) \times (3^2)$			1			1						$(5,1) \times (2^3)$
$(5,1) \times (3,2,1)$			1	1	1	2	1	1	1			$(5,1) \times (3,2,1)$
$(4,2) \times (4,2)$	1	1	2	1		2	1	1				$(4,2) \times (2^2,1^2)$
$(4,2) \times (4,1^2)$		1	1	2	1	2	1		1			$(4,2) \times (3,1^3)$
$(4,2) \times (3^2)$		1		1	1	1			1			$(4,2) \times (2^3)$
$(4,2) \times (3,2,1)$		1	2	2	1	3	2	1	2	1		$(4,2) \times (3,2,1)$
$(4,1^2) \times (4,1^2)$	1	1	2	1	1	2	1	1	1	1		$(4,1^2) \times (3,1^3)$
$(4,1^2) \times (3^2)$			1	1		1	1	1				$(4,1^2) \times (2^3)$
$(4,1^2) \times (3,2,1)$		1	2	2	1	4	2	1	2	1		$(4,1^2) \times (3,2,1)$
$(3^2) \times (3^2)$	1		1				1	1				$(3^2) \times (2^3)$
$(3^2) \times (3,2,1)$		1	1	1		2	1		1	1		$(3^2) \times (3,2,1)$
$(3,2,1) \times (3,2,1)$	1	2	3	4	2	5	4	2	3	2	1	$(3,2,1) \times (3,2,1)$
	(1^6)	$(2,1^4)$	$(2^2,1^2)$	$(3,1^3)$	(2^3)	$(3,2,1)$	$(4,1^2)$	(3^2)	$(4,2)$	$(5,1)$	(6)	

5. $n=7$.

	(7)	$(6,1)$	$(5,2)$	$(5,1^2)$	$(4,3)$	$(4,2,1)$	$(4,1^3)$	$(3^2,1)$	$(3,2^2)$	$(3,2,1^2)$	$(3,1^4)$	$(2^3,1)$	$(2^2,1^3)$	$(2,1^5)$	(1^7)	
$(6,1)\times(6,1)$	1	1	1	1												$(6,1)\times(2,1^5)$
$(6,1)\times(5,2)$		1	1	1	1	1										$(6,1)\times(2^2,1^3)$
$(6,1)\times(5,1^2)$		1	1	1		1	1									$(6,1)\times(3,1^4)$
$(6,1)\times(4,3)$			1		1	1		1								$(6,1)\times(2^3,1)$
$(6,1)\times(4,2,1)$			1	1	1	2	1	1	1	1						$(6,1)\times(3,2,1^2)$
$(6,1)\times(4,1^3)$				1		1	1			1	1					$(6,1)\times(4,1^3)$
$(6,1)\times(3^2,1)$					1	1		1	1	1						$(6,1)\times(3,2^2)$
$(5,2)\times(5,2)$	1	1	2	1	1	2	1	1	1							$(5,2)\times(2^2,1^3)$
$(5,2)\times(5,1^2)$		1	1	2	1	2	1	1		1						$(5,2)\times(3,1^4)$
$(5,2)\times(4,3)$		1	1	1	1	2		1	1	1						$(5,2)\times(2^3,1)$
$(5,2)\times(4,2,1)$		1	2	2	2	4	2	2	2	3	1	1				$(5,2)\times(3,2,1^2)$
$(5,2)\times(4,1^3)$			1	1		2	2	1	1	2	1		1			$(5,2)\times(4,1^3)$
$(5,2)\times(3^2,1)$			1	1	1	2	1	2	1	2		1	1			$(5,2)\times(3,2^2)$
$(5,1^2)\times(5,1^2)$	1	1	2	1	1	2	1		1	1	1					$(5,1^2)\times(3,1^4)$
$(5,1^2)\times(4,3)$			1	1	1	2	1	1	1	1						$(5,1^2)\times(2^3,1)$
$(5,1^2)\times(4,2,1)$		1	2	2	2	4	2	3	2	3	1	1	1			$(5,1^2)\times(3,2,1^2)$
$(5,1^2)\times(4,1^3)$		1	1	1	1	2	1	1	1	2	1	1	1	1		$(5,1^2)\times(4,1^3)$
$(5,1^2)\times(3^2,1)$			1		1	3	1	1	2	2	1	1				$(5,1^2)\times(3,2^2)$
$(4,3)\times(4,3)$	1	1	1	1	1	1	1	1	1	1		1				$(4,3)\times(2^3,1)$
$(4,3)\times(4,2,1)$		1	2	2	1	4	2	2	2	3	1	1	1			$(4,3)\times(3,2,1^2)$
$(4,3)\times(4,1^3)$				1	1	2	2	1	1	2	1	1				$(4,3)\times(4,1^3)$
$(4,3)\times(3^2,1)$		1	1	1	1	2	1	1	1	2	1	1	1			$(4,3)\times(3,2^2)$
$(4,2,1)\times(4,2,1)$	1	2	4	4	4	9	5	5	5	8	3	3	3	1		$(4,2,1)\times(3,2,1^2)$
$(4,2,1)\times(4,1^3)$		1	2	2	2	5	2	3	3	5	2	2	2	1		$(4,2,1)\times(4,1^3)$
$(4,2,1)\times(3^2,1)$		1	2	3	2	5	3	3	3	5	2	2	2	1		$(4,2,1)\times(3,2^2)$
$(4,1^3)\times(4,1^3)$	1	1	2	1	2	2	1	2	2	2	1	2	2	1		$(4,1^3)\times(4,1^3)$
$(4,1^3)\times(3^2,1)$			1	1	1	3	2	2	2	3	1	1	1			$(4,1^3)\times(3,2^2)$
$(3^2,1)\times(3^2,1)$	1	1	2	1	1	3	2	1	2	3	2	1	1	1		$(3^2,1)\times(3,2^2)$
	(1^7)	$(2,1^5)$	$(2^2,1^3)$	$(3,1^4)$	$(2^3,1)$	$(3,2,1^2)$	$(4,1^3)$	$(3,2^2)$	$(3^2,1)$	$(4,2,1)$	$(5,1^2)$	$(4,3)$	$(5,2)$	$(6,1)$	(7)	

6. $n=8$.

	(8)	$(7,1)$	$(6,2)$	$(6,1^2)$	$(5,3)$	$(5,2,1)$	$(5,1^3)$	(4^2)	$(4,3,1)$	$(4,2^2)$	$(4,2,1^2)$	$(4,1^4)$	$(3^2,2)$	$(3^2,1^2)$	$(3,2^2,1)$	$(3,2,1^3)$	$(3,1^5)$	(2^4)	$(2^3,1^2)$	$(2^2,1^4)$	$(2,1^6)$	(1^8)	
$(7,1) \times (7,1)$	1	1	1	1																			$(7,1) \times (2,1^6)$
$(7,1) \times (6,2)$		1	1	1	1	1																	$(7,1) \times (2^2,1^4)$
$(7,1) \times (6,1^2)$		1	1	1		1	1																$(7,1) \times (3,1^5)$
$(7,1) \times (5,3)$			1		1	1		1	1														$(7,1) \times (2^3,1^2)$
$(7,1) \times (5,2,1)$			1	1	1	2	1		1	1	1												$(7,1) \times (3,2,1^3)$
$(7,1) \times (5,1^3)$				1		1	1				1	1											$(7,1) \times (4,1^4)$
$(7,1) \times (4^2)$					1				1														$(7,1) \times (2^4)$
$(7,1) \times (4,3,1)$					1	1		1	2	1	1		1	1									$(7,1) \times (3,2^2,1)$
$(7,1) \times (4,2^2)$						1			1	1	1		1		1								$(7,1) \times (3^2,1^2)$
$(7,1) \times (4,2,1^2)$						1	1		1	1	2	1		1	1	1							$(7,1) \times (4,2,1^2)$
$(7,1) \times (3^2,2)$									1	1			1	1	1								$(7,1) \times (3^2,2)$
$(6,2) \times (6,2)$	1	1	2	1	1	2	1	1	1	1													$(6,2) \times (2^2,1^4)$
$(6,2) \times (6,1^2)$		1	1	2	1	2	1		1		1												$(6,2) \times (3,1^5)$
$(6,2) \times (5,3)$		1	1	1	2	2			2	1	1		1										$(6,2) \times (2^3,1^2)$
$(6,2) \times (5,2,1)$		1	2	2	2	4	2	1	3	2	3	1	1	1	1								$(6,2) \times (3,2,1^3)$
$(6,2) \times (5,1^3)$			1	1		2	2		1	1	2	1		1		1							$(6,2) \times (4,1^4)$
$(6,2) \times (4^2)$			1			1		1	1	1				1									$(6,2) \times (2^4)$
$(6,2) \times (4,3,1)$			1	1	2	3	1	1	4	2	3		2	2	2	1							$(6,2) \times (3,2^2,1)$
$(6,2) \times (4,2^2)$			1		1	2	1	1	2	3	2	1	1	2	2	1		1					$(6,2) \times (3^2,1^2)$
$(6,2) \times (4,2,1^2)$				1	1	3	2		3	2	5	2	2	2	3	3	1		1				$(6,2) \times (4,2,1^2)$
$(6,2) \times (3^2,2)$					1	1			2	1	2		2	1	2	1			1				$(6,2) \times (3^2,2)$
$(6,1^2) \times (6,1^2)$	1	1	2	1	1	2	1			1	1	1											$(6,1^2) \times (3,1^5)$
	(1^8)	$(2,1^6)$	$(2^2,1^4)$	$(3,1^5)$	$(2^3,1^2)$	$(3,2,1^3)$	$(4,1^4)$	(2^4)	$(3,2^2,1)$	$(3^2,1^2)$	$(4,2,1^2)$	$(5,1^3)$	$(3^2,2)$	$(4,2^2)$	$(4,3,1)$	$(5,2,1)$	$(6,1^2)$	(4^2)	$(5,3)$	$(6,2)$	$(7,1)$	(8)	

↑

	(8)	$(7, 1)$	$(6, 2)$	$(6, 1^2)$	$(5, 3)$	$(5, 2, 1)$	$(5, 1^3)$	(4^2)	$(4, 3, 1)$	$(4, 2^2)$	$(4, 2, 1^2)$	$(4, 1^4)$	$(3^2, 2)$	$(3^2, 1^2)$	$(3, 2^2, 1)$	$(3, 2, 1^3)$	$(3, 1^5)$	(2^4)	$(2^3, 1^2)$	$(2^2, 1^4)$	$(2, 1^6)$	(1^8)	
$(6, 1^2) \times (5, 3)$			1	1	1	2	1	1	2	1	1			1									$(6, 1^2) \times (2^3, 1^2)$
$(6, 1^2) \times (5, 2, 1)$		1	2	2	2	4	2	1	3	2	3	1	1	1	1	1							$(6, 1^2) \times (3, 2, 1^3)$
$(6, 1^2) \times (5, 1^3)$		1	1	1	1	2	1		1	1	2	1			1	1	1						$(6, 1^2) \times (4, 1^4)$
$(6, 1^2) \times (4^2)$					1	1			1		1		1										$(6, 1^2) \times (2^4)$
$(6, 1^2) \times (4, 3, 1)$			1		2	3	1	1	4	3	3	1	2	2	2	1							$(6, 1^2) \times (3, 2^2, 1)$
$(6, 1^2) \times (4, 2^2)$				1	1	2	1		3	1	3		2	2	2	1			1				$(6, 1^2) \times (3^2, 1^2)$
$(6, 1^2) \times (4, 2, 1^2)$			1	1	1	3	2	1	3	3	4	2	2	3	3	3	1	1	1	1			$(6, 1^2) \times (4, 2, 1^2)$
$(6, 1^2) \times (3^2, 2)$						1		1	2	2	2		1	2	2	1		1					$(6, 1^2) \times (3^2, 2)$
$(5, 3) \times (5, 3)$	1	1	2	1	1	2	1	1	2	2	1		1	1	1								$(5, 3) \times (2^3, 1^2)$
$(5, 3) \times (5, 2, 1)$		1	2	2	2	5	2	1	4	3	4	1	2	2	2	1							$(5, 3) \times (3, 2, 1^3)$
$(5, 3) \times (5, 1^3)$				1	1	2	2		2	1	3	1	1	1	1	1							$(5, 3) \times (4, 1^4)$
$(5, 3) \times (4^2)$		1		1	1	1			1		1		1		1								$(5, 3) \times (2^4)$
$(5, 3) \times (4, 3, 1)$		1	2	2	2	4	2	1	4	3	4	1	2	3	3	2		1	1				$(5, 3) \times (3, 2^2, 1)$
$(5, 3) \times (4, 2^2)$			1	1	2	3	1		3	2	4	1	2	1	3	2	1		1				$(5, 3) \times (3^2, 1^2)$
$(5, 3) \times (4, 2, 1^2)$			1	1	1	4	3	1	4	4	6	3	2	4	4	4	1	1	1	1			$(5, 3) \times (4, 2^2, 1)$
$(5, 3) \times (3^2, 2)$			1		1	2	1	1	2	2	2	1	1	2	2	2		1	1	1			$(5, 3) \times (3^2, 2)$
$(5, 2, 1) \times (5, 2, 1)$	1	2	4	4	5	9	5	2	9	6	9	3	4	5	5	4	1	1	1				$(5, 2, 1) \times (3, 2, 1^3)$
$(5, 2, 1) \times (5, 1^3)$		1	2	2	2	5	2	1	4	3	5	2	2	3	3	3	1		1	1			$(5, 2, 1) \times (4, 1^4)$
$(5, 2, 1) \times (4^2)$			1	1	1	2	1		2	2	2		1	1	1	1							$(5, 2, 1) \times (2^4)$
$(5, 2, 1) \times (4, 3, 1)$		1	3	3	4	9	4	2	9	7	10	3	5	6	7	5	1	1	2	1			$(5, 2, 1) \times (3, 2^2, 1)$
$(5, 2, 1) \times (4, 2^2)$		1	2	2	3	6	3	2	7	5	8	3	4	5	6	5	1	1	2	1			$(5, 2, 1) \times (3^2, 1^2)$
$(5, 2, 1) \times (4, 2, 1^2)$		1	3	3	4	9	5	2	10	8	13	5	6	8	10	9	3	2	4	3	1		$(5, 2, 1) \times (4, 2, 1^2)$
	(1^8)	$(2, 1^6)$	$(2^2, 1^4)$	$(3, 1^5)$	$(2^3, 1^2)$	$(3, 2, 1^3)$	$(4, 1^4)$	(2^4)	$(3, 2^2, 1)$	$(3^2, 1^2)$	$(4, 2, 1^2)$	$(5, 1^3)$	$(3^2, 2)$	$(4, 2^2)$	$(4, 3, 1)$	$(5, 2, 1)$	$(6, 1^2)$	(4^2)	$(5, 3)$	$(6, 2)$	$(7, 1)$	(8)	

↓

	(8)	$(7, 1)$	$(6, 2)$	$(6, 1^2)$	$(5, 3)$	$(5, 2, 1)$	$(5, 1^3)$	(4^2)	$(4, 3, 1)$	$(4, 2^2)$	$(4, 2, 1^2)$	$(4, 1^4)$	$(3^2, 2)$	$(3^2, 1^2)$	$(3, 2^2, 1)$	$(3, 2, 1^3)$	$(3, 1^5)$	(2^4)	$(2^3, 1^2)$	$(2^2, 1^4)$	$(2, 1^6)$	(1^8)	
$(5, 2, 1) \times (3^2, 2)$			1	1	2	4	2	1	5	4	6	2	3	4	5	4	1	1	2	1			$(5, 2, 1) \times (3^2, 2)$
$(5, 1^3) \times (5, 1^3)$	1	1	2	1	2	2	1	1	2	2	2	1	1	1	2	2	1	1	1	1	1		$(5, 1^3) \times (4, 1^4)$
$(5, 1^3) \times (4^2)$						1	1	1	1	1	1	1		1	1								$(5, 1^3) \times (2^4)$
$(5, 1^3) \times (4, 3, 1)$			1	1	2	4	2	1	5	4	6	2	3	3	4	3	1	1	1				$(5, 1^3) \times (3, 2^2, 1)$
$(5, 1^3) \times (4, 2^2)$			1	1	1	3	2	1	4	3	4	1	2	4	3	3		1	1	1			$(5, 1^3) \times (3^2, 1^2)$
$(5, 1^3) \times (4, 2, 1^2)$		1	2	2	3	5	2	1	6	4	6	2	4	4	6	5	2	1	3	2	1		$(5, 1^3) \times (4, 2, 1^2)$
$(5, 1^3) \times (3^2, 2)$					1	2	1		3	2	4	1	2	2	3	2			1				$(5, 1^3) \times (3^2, 2)$
$(4^2) \times (4^2)$	1		1				1	1		1				1				1					$(4^2) \times (2^4)$
$(4^2) \times (4, 3, 1)$		1	1	1	1	2	1		2	1	2	1	1	1	2	1			1				$(4^2) \times (3, 2^2, 1)$
$(4^2) \times (4, 2^2)$			1			2	1	1	1	2	1	1		2	1	1		1		1			$(4^2) \times (3^2, 1^2)$
$(4^2) \times (4, 2, 1^2)$				1	1	2	1		2	1	4	1	2	1	2	2	1		1				$(4^2) \times (4, 2, 1^2)$
$(4^2) \times (3^2, 2)$				1	1	1			1		2		1		1	1	1		1				$(4^2) \times (3^2, 2)$
$(4, 3, 1) \times (4, 3, 1)$	1	2	4	4	4	9	5	2	8	7	11	4	5	6	8	7	2	2	3	2			$(4, 3, 1) \times (3, 2^2, 1)$
$(4, 3, 1) \times (4, 2^2)$		1	2	3	3	7	4	1	7	5	9	3	4	5	6	6	2	1	3	2	1		$(4, 3, 1) \times (3^2, 1^2)$
$(4, 3, 1) \times (4, 2, 1^2)$		1	3	3	4	10	6	2	11	9	14	6	6	9	11	10	3	2	4	3	1		$(4, 3, 1) \times (4, 2, 1^2)$
$(4, 3, 1) \times (3^2, 2)$		1	2	2	2	5	3	1	5	4	6	3	2	4	5	5	2	1	2	2	1		$(4, 3, 1) \times (3^2, 2)$
$(4, 2^2) \times (4, 2^2)$	1	1	3	1	2	5	3	2	5	6	6	4	2	5	5	5	2	2	1	2			$(4, 2^2) \times (3^2, 1^2)$
$(4, 2^2) \times (4, 2, 1^2)$		1	2	3	4	8	4	1	9	6	12	4	6	6	9	8	3	1	4	2	1		$(4, 2^2) \times (4, 2, 1^2)$
$(4, 2^2) \times (3^2, 2)$		1	1	2	2	4	2		4	2	6	2	3	2	4	4	2		2	1	1		$(4, 2^2) \times (3^2, 2)$
$(4, 2, 1^2) \times (4, 2, 1^2)$	1	2	5	4	6	13	6	4	14	12	17	6	8	12	14	13	4	4	6	5	2	1	$(4, 2, 1^2) \times (4, 2, 1^2)$
$(4, 2, 1^2) \times (3^2, 2)$			2	2	2	6	4	2	6	6	8	4	4	6	6	6	2	2	2	2			$(4, 2, 1^2) \times (3, 2^2)$
$(3^2, 2) \times (3^2, 2)$	1	1	2	1	1	3	2	1	2	3	4	2	1	3	2	3	1	1	1	2	1	1	$(3^2, 2) \times (3^2, 2)$
	(1^8)	$(2, 1^6)$	$(2^2, 1^4)$	$(3, 1^5)$	$(2^3, 1^2)$	$(3, 2, 1^3)$	$(4, 1^4)$	(2^4)	$(3, 2^2, 1)$	$(3^2, 1^2)$	$(4, 2, 1^2)$	$(5, 1^3)$	$(3^2, 2)$	$(4, 2^2)$	$(4, 3, 1)$	$(5, 2, 1)$	$(6, 1^2)$	(4^2)	$(5, 3)$	$(6, 2)$	$(7, 1)$	(8)	

REFERENCES.

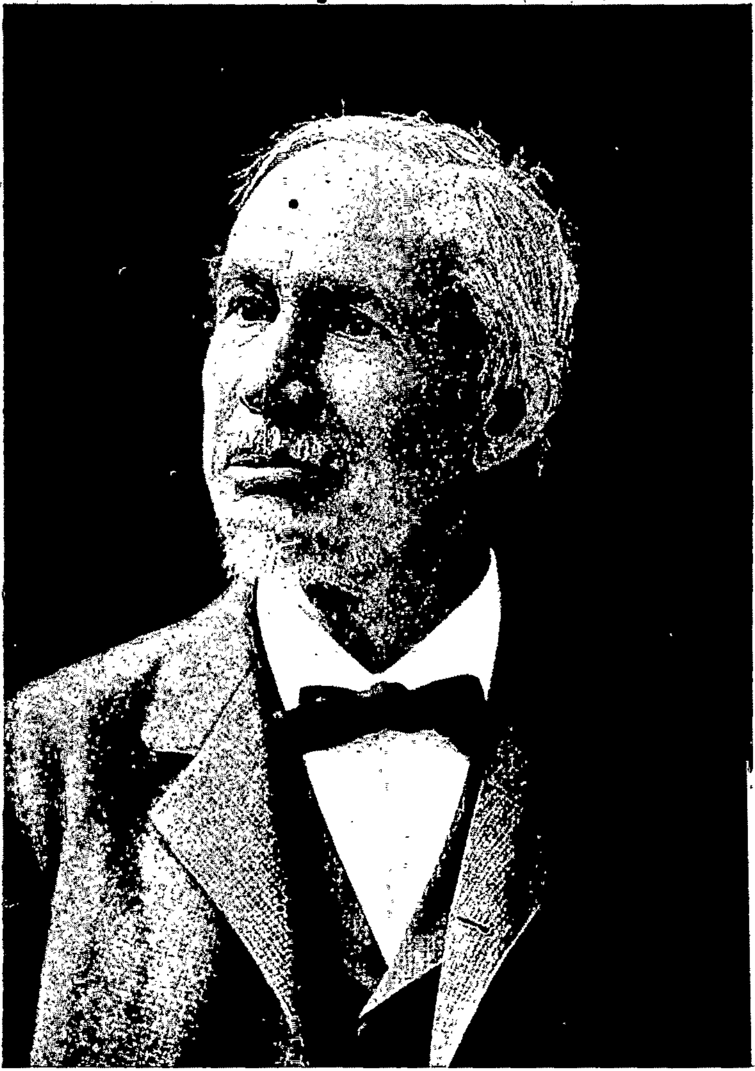
1. R. Brauer, "Sur la multiplication des caractéristiques des groupes continus et semi-simples," *Comptes rendus*, **204** (1937), pp. 1784-1786.
2. F. D. Murnaghan, "On the representations of the symmetric group," *American Journal of Mathematics*, **59** (1937), pp. 437-488.
3. ———, "The analysis of the direct product of irreducible representations of the symmetric groups," *American Journal of Mathematics*, **60** (1938), pp. 44-65.
4. ———, "The characters of the symmetric group," *American Journal of Mathematics*, **59** (1937), pp. 739-753.
5. D. E. Littlewood, "Group characters and the structure of groups," *Proceedings of the London Mathematical Society*, Ser. 2, **39** (1935), pp. 150-204.

THE JOHNS HOPKINS UNIVERSITY.

CALCUTTA.
UNIVERSITY.
CENTRAL
LIBRARY.

Memorial Number

George William Hill



G. W. Hill.

CONTENTS

PAGE

The equations of motion of the Moon. By E. W. Brown, ... 785

Sur les avances du noeud et du périhélie d'une planète sous l'action d'une anneau circulaire. By Jean Chazy, ... 793

Die symmetrischen periodischen Bahnen des restringierten Dreikörperproblems in der Nachbarschaft eines kritischen Keplerkreises. By Ernst Hölder, ... 801

Symbolic dynamics. By Marston Morse and Gustav A. Hedlund, . 815

Periodic orbits in the restricted problem of three bodies in their relation to Hill's work on the motion of the Moon. By Elis Strömgren, 867

Zur Theorie der Dirichletschen Reihen. By Otto Toeplitz, . . 880

Mean motion. By Hermann Weyl, ... 889

The homogeneous chaos. By Norbert Wiener, ... 897

On Hill's periodic lunar orbit. By Aurel Wintner, ... 937

THE EQUATIONS OF MOTION OF THE MOON.

By Ernest W. Brown.[*]

1. In Volume 1 of this Journal, G. W. Hill published a paper entitled "Researches in the Lunar Theory" which has had far-reaching effects on the later development of the subject. From a study of it H. Poincaré conceived and developed the idea of the periodic orbit which in his hands became the basis for an extended study of the mathematics of the problem of three bodies. The calculation and classification of orbits of the type studied by Hill have formed the subject of many papers by various authors since his time. In my own case, it formed the starting point for a complete recalculation of the action of the sun on the motion of the moon. It is mainly the methods used in this last work that I am going to reconsider here.

It is not wise, I believe, to imagine that the work of a great master in his subject cannot be improved, especially when extended calculations are involved. As time goes on, new techniques are developed and in our own time calculating machines have come within the reach of everyone, so that a method previously rejected on account of the extended computation required may become easy and simple with but slight changes. My own experience in several cases leads me to think that it is worth while to return to the foundations even when they go so far back in time as those of Celestial Mechanics, and to examine whether valuable changes cannot be made.

2. The present paper is the result of such an examination. In my calculation, I adopted Hill's equations as they stood merely adding the terms in the disturbing function which complete accuracy demanded and adapting the solution of the equations to the additional terms. Two or more years ago it occurred to me to try the effect of referring the motion to axes which followed the mean position of the moon. I found later that Euler had had the same idea in his second theory. Hill had used axes which followed the mean position of the sun. There was a reason for this change as the coördinates would then take the form $a + x$, y where x, y are small and therefore expansions in powers of x, y were possible. What I did not expect, however, was that the equations for x, y could be put into a form such that the first gave x and then the second gave y with the same degree of approximation that had

---

[*] Professor Brown, who successfully undertook the great task of developing and applying Hill's theory of the Moon's motion, died on July 22, 1938, some days after revising the proof of the present paper. Editors.

been obtained before by the troublesome solution of two simultaneous equations. Further, the homogenous equation for x had fewer large terms present than the earlier equation which it superseded, that for y being substantially the same as before.

A further advantage in calculation is due to the fact that the lowest power of n' or m present is the second; in Hill's equations the first power occurs. Finally they have permitted me to find out the loss of accuracy due to small divisors which has occurred through the whole theory. This last feature is developed below and leads to the satisfactory result that the loss is considerably less than I had supposed. This fact has already appeared in the verification now under way.

When $e' = 0$, the time disappears explicitly from Hill's equations. This property has considerable advantage in theoretical researches and in the search for periodic orbits. But it is of no advantage in the calculation of satellite orbits. On the contrary it is a disadvantage in the latter case because the elimination of t deprives the equations of the advantages mentioned above for the new form.

3. The next step was an attempt to abbreviate the calculation of the various terms in the equations. In the so-called homogeneous form, if we neglect the small ratio of the distances, every term is a second degree term in x, y, z or their derivatives. As the coördinates consist of series of harmonic terms, by far the greatest part of the work consists of the multiplication of pairs of harmonic series. If then a technique could be developed for the multiplication of harmonic series by machinery practically the whole of the work of calculating the action of the sun on the moon could be done in perhaps a tenth of the time it originally required. I was fortunate in interesting my friend and former pupil, Professor W. J. Eckert of Columbia University, in this problem. He had already adapted several commercial machines of the Hollerith type to the solution of problems of celestial mechanics, and had developed a computing laboratory for their effective use. With the aid of the facilities of this laboratory we have been able in a year or two to test and extend calculations which took me nearly twenty years to carry out with the old-fashioned methods. These facts are mentioned because we are using the new equations in this verification.

4. The equations have also been used in another connection, namely, the exact calculation[1] of a portion of the motion of the perigee which had been neglected as insensible in the earlier work, but which I had found by another method[2] must be large enough to account for most of the previously unexplained difference between the observed and theoretical values.

---

[1] *Monthly Notices of the Royal Astronomical Society*, vol. 97 (1936), pp. 116-127.

These equations appear to be the best to adopt for satellite motion when we can use the variation orbit as a first approximation and expand in powers of the eccentricities, inclination and ratio of the distances. For a satellite with a large eccentricity like Jupiter VIII which seems to require the ellipse as a first approximation, I hope to give elsewhere a method which appears to have considerable advantages over that which Professor Brouwer and I have used for this body and for which I was responsible.

The Non-Homogeneous Equations.

5. The equations of motion referred to an x-axis in the direction of the mean position of the moon, longitude $nt + \epsilon$, the y-axis perpendicular to it in a fixed plane and the z-axis perpendicular to this plane are:

$$5.1 \qquad \left\{\begin{array}{llll} \ddot{x} - 2n\dot{y} + & (1+x)\left(\dfrac{\mu}{\mathrm{a}^3 r^3} - n^2\right) & = \dfrac{\partial \Omega}{\partial x}, \\ \ddot{y} + 2n\dot{x} + & y\left(\dfrac{\mu}{\mathrm{a}^3 r^3} - n^2\right) & = \dfrac{\partial \Omega}{\partial y}, \\ \ddot{z} \qquad\quad + & z \cdot \dfrac{\mu}{\mathrm{a}^3 r^3} & = \dfrac{\partial \Omega}{\partial z}, \end{array}\right.$$

where $\mathrm{a}(1 + x)$, $\mathrm{a}y$, $\mathrm{a}z$ are the coördinates of the moon, $\mathrm{a}r$ the distance and $\Omega \mathrm{a}^2$ the disturbing function due to the sun, supposed to move in an ellipse in the xy-plane.

When the moon is undisturbed and moves in a circular orbit in the plane of xy we have

$$x = y = z = 0, \qquad r = 1, \qquad \mathrm{a} = a, \qquad \mu/\mathrm{a}^3 = n^2.$$

Hence the factor in brackets in the first two equations is small.

If we add the integral of the second equation multiplied by $2n$ to the first we obtain an equation which may be written

$$\ddot{x} + (1+x)\left(\frac{\mu}{\mathrm{a}^3 r^3} + 3n^2\right) = \frac{\partial \Omega}{\partial x} + 2n \int \left\{ \frac{\partial \Omega}{\partial y} - y\left(\frac{\mu}{\mathrm{a}^3 r^3} - n^2\right) \right\} dt + \text{const.}$$

Since

$$1/r^3 = 1 - 3x + 6x^2 - \tfrac{3}{2}y^2 - \tfrac{3}{2}z^2 + \cdots$$

$\mu/\mathrm{a}^3 r^3$ differs from n^2 by a small quantity of at least the first order and we know that Ω contains a factor of the second order; this equation therefore gives.

$$\ddot{x} + n^2 x = \text{terms of the second order} + \text{const.}$$

---

[2] *Ibid.*, vol. 97 (1936), pp. 56-66

provided the integral introduces no small divisors of the first order. The point of this equation is the fact that y or z are not present as first order terms. The equation can therefore be solved to give an approximation to x.

Since the last term on the left is of the second order, the second equation of 5.1 evidently gives y when we know x. The third equation gives z in the same manner as the first gives x.

6. The actual procedure is so similar to that used in my Theory of the Motion of the Moon, that it is not necessary to repeat the various steps here. The chief difference is due to the fact that the various approximations to x, y are obtained separately instead of together by the solution of a pair of simultaneous equations as in the work mentioned. It may be noted that the $a(1+x)+aiy$ of this paper is the same as the $u\zeta^{-1}$ of the earlier work.

As the form of Ω differs somewhat I shall give it completely below, as far as it is needed.

We have[3]

$$6.1 \qquad \Omega \mathrm{a}^2 = \frac{m'\mathrm{a}^2}{a'^3}\left(\omega_2 + \frac{E-M}{E+M}\omega_3 + \frac{E^2-EM+M^2}{(E+M)^2}\omega_4 + \cdots\right),$$

where

$$\omega_2 = \frac{a'^3}{r'^3}\left(\tfrac{3}{2}r^2S^2 - r^2/2\right), \qquad \omega_3 = \frac{a'^4}{r'^4}\,\frac{\mathrm{a}}{\mathrm{a}'}\left(\tfrac{5}{2}r^3S^3 - \tfrac{3}{2}r^3S\right),$$

$$\omega_4 = \frac{a'^5}{r'^5}\,\frac{\mathrm{a}^2}{\mathrm{a}'^2}\left(\tfrac{35}{8}S^4r^4 - \tfrac{15}{4}S^2r^4 + \tfrac{3}{8}r^4\right),$$

$$rr'S = (1+x)r'\cos(n't+\epsilon'-nt-\epsilon+E') + yr'\sin(n't+\epsilon'-nt-\epsilon+E'),$$

$$r^2 = (1+x)^2+y^2+z^2,$$

where $n't+\epsilon'$, E' are the mean longitude and equation of the center of the sun's orbit. In 6.1, E, M are the masses of the earth and moon.

In order to expand these in powers of e' it will be found convenient to use complex quantities by putting

$$1+x+iy=u_1, \quad 1+x-iy=s_1, \quad nt+\epsilon-n't-\epsilon'=D, \quad e^{Di}=\zeta;$$

E', r' are functions of $l'=n't+\epsilon'-\omega'$, an angle which disappears when $e'=0$. The expansions of the various functions of r', E' needed will then be found in T. M. M.[4]

We also put

$$m'+E+M=n'^2a'^3, \quad \frac{m'}{a'^3n^2}=\frac{n'^2}{n^2}(1-q)=m^2(1-q), \quad q=\frac{E+M}{E+M+m'}.$$

---

[3] Brown, *Lunar Theory*, Chap. I.

[4] These letters are used to denote my "Theory of the motion of the Moon," published in vols. 53 and 59 of the *Memoirs of the Royal Astronomical Society.*

If we are using numerical values, that of q can be inserted at the outset. Otherwise we put $q = 0$ and make a correction for it at the end.

THE HOMOGENEOUS EQUATIONS.

7. Multiply the equations 5.1 by $\dot{x}$, $\dot{y}$, $\dot{z}$ respectively, add and integrate. Noticing that t occurs in Ω, through the coördinates x, y, z as well as explicitly where it is shown and where it is present in E', r' and denoting this explicit occurrence by $\partial\Omega/\partial t$, we have

$$7.1 \quad \tfrac{1}{2}(\dot{x}^2 + \dot{y}^2 + \dot{z}^2) - \frac{\mu}{a^3 r} - \tfrac{1}{2}n^2\{(1 + x)^2 + y^2\} = \Omega - \int \frac{\partial\Omega}{\partial t}\, dt + \text{const.}$$

Now t occurs explicitly through $D = nt + \epsilon - n't - \epsilon'$ and through $n't + \epsilon' - \omega' = l'$. We have therefore

$$\frac{\partial\Omega}{\partial t} = \frac{\partial\Omega}{\partial S}\cdot\frac{\partial S}{\partial D}(n - n') + \frac{\partial\Omega}{\partial l'}n'$$
$$= \frac{\partial\Omega}{\partial S}\left\{ -\frac{1 + x}{r}\sin(D - E') - \frac{y}{r}\cos(D - E') \right\}(n - n') + \frac{\partial\Omega}{\partial l'}n'.$$

But

$$(1 + x)\frac{\partial\Omega}{\partial y} - y\frac{\partial\Omega}{\partial x} = \frac{\partial\Omega}{\partial S}\left\{ -\frac{1 + x}{r}\sin(D - E') - \frac{y}{r}\cos(D - E') \right\}.$$

Hence

$$7.2 \quad \frac{\partial\Omega}{\partial t} = \left\{ (1 + x)\frac{\partial\Omega}{\partial y} - y\frac{\partial\Omega}{\partial x} \right\}(n - n') + \frac{\partial\Omega}{\partial l'}n' = \frac{dQ}{dt}(n - n') + \frac{\partial\Omega}{\partial l'}n'$$

this giving the definition of Q. The right-hand member of 7.1 is therefore

$$\Omega - Q(n - n') - n'\int \frac{\partial\Omega}{\partial l'}\, dt + \text{const.}$$

The integral vanishes with e'.

Again, multiply equations 5.1 by $1 + x$, y, z respectively and add; we obtain a result which may be written

$$7.3 \quad \tfrac{1}{2}\frac{d^2}{dt^2}(r^2) - (\dot{x}^2 + \dot{y}^2 + \dot{z}^2) - 2n\{\dot{y}(1 + x) - \dot{x}y\}$$
$$+ \frac{\mu}{a^3 r} - n^2\{(1 + x)^2 + y^2\} = r\frac{\partial\Omega}{\partial r},$$

where Ω in the right-hand member is supposed to be expressed in terms of r, S.

Adding this to 7.1, we note that the negative power of r disappears and we get

$$7.4 \quad \tfrac{1}{2}\frac{d^2}{dt^2}(r^2) - \tfrac{1}{2}(\dot{x}^2 + \dot{y}^2 + \dot{z}^2) - 2n\{\dot{y}(1 + x) - \dot{x}y\} - \tfrac{3}{2}n^2\{(1 + x)^2 + y^2\}$$
$$= \Omega + r\frac{\partial\Omega}{\partial r} - Q(n - n') - n'\int \frac{\partial\Omega}{\partial l'}\, dt + \text{const.}$$

By multiplying the first two equations of 5. 1 by $-y$, $1+x$, and adding, we get a second equation independent of negative powers of r. This may be written after integration

$$7.5 \qquad \dot{y}(1+x)-\dot{x}y+n\{(1+x)^2+y^2\}=Q+\text{const.}$$

Finally, multiplying this by $2n$ and adding the result to the previous equation we obtain

$$7.6 \qquad \tfrac{1}{2}\frac{d^2}{dt^2}(r^2)+\tfrac{1}{2}n^2\{(1+x)^2+y^2\}-\tfrac{1}{2}(\dot{x}^2+\dot{y}^2+\dot{z}^2)$$
$$=\Omega+r\frac{\partial\Omega}{\partial r}+Q(n+n')-n'\int\frac{\partial\Omega}{\partial l'}\,dt+\text{const.}$$

For calculation subtract from each side of the equation $-\frac{3}{4}m^2(x^2+y^2-2z^2)$. The right-hand member will then contain no term free from one or both of the explicit angles $2D$, l' except a small term depending on the square of the ratio of the parallaxes.

To get an equation for z, we subtract the first of 5. 1 multiplied by z from the third multiplied by $(1+x)$ and obtain

$$7.7 \qquad (1+x)(\ddot{z}+n^2z)+z\left\{-\ddot{x}+2n\dot{y}-(1+x)\frac{\partial\Omega}{zdz}+\frac{\partial\Omega}{\partial x}\right\}=0.$$

The equations 7. 6, 7. 5, 7. 7 are those needed. It will be noticed that if we express the coördinate $1+x$ by a single letter and confine Ω to the portion ω_2, the equations are homogeneous and of the second degree in x, y, z apart from the arbitrary constants.

8. Constants. The constants of the theory are found as in T. M. M., except perhaps the constant a. It is best to leave this to the end of the work and find it from the first of equations 5. 1, or 7. 3, the latter having the advantage of also giving $1/r$. In substituting the results obtained for x, y, z therein, I recall that here x is defined to contain no constant term. It consists entirely of cosines of multiples of the angles D, l, l', $2F$, while y consists of sines of the same angles. In the coördinate z the multiples of F present are odd, and the harmonics are sines.

9. Loss of accuracy due to small divisors. The approximate parameters are $m=1/13$, $e=1/20$, $\gamma=\sin\frac{1}{2}i=1/22$, $e'=1/60$, $a/a'=1/400$. The effects of the first three of these generally will be discussed as though each of them was 1/10.

Small divisors may arise in the x, z equations from terms which have a period of a month, and in the y equation from terms of periods long compared with a month. The latter will also arise in Q so that y may involve the square of the small divisor.

Small divisors are of two kinds. Both of them are independent of the mean motion of the moon. Those containing the mean motion of the sun will have the factor m and those independent of it the factor m^2. Examples of the former are $2D-l$, D, l', $2D-2l$, and of the latter, $2F-2l$, $D+l'$, $D+l'-l$. It will be noticed that each long period term is associated with two short period terms obtained by adding and subtracting l from its argument.

Suppose that we determine at any one step, as in T. M. M., the terms containing a given characteristic, i. e., those which have as a factor any given powers of e, γ, e', a/a'. In any such group there are either monthly or long period terms but not both. This is because the arguments in any such group always differ by even multiples of D, l, l', F. Thus as we proceed from lower to higher characteristics, short period and long period terms occur alternately, except as far as multiples of l' are concerned. But as $e'=1/60$, and the main argument proceeds by decimal places, the additional factor 1/6 takes care of accumulations and no special treatment is necessary.

Let us first take the cases of divisors which have m as a factor. Since Ω contains the explicit factor m^2 there is no loss of accuracy from these terms in Ω even when a double integration is involved. Thus the only terms which might cause a loss are those in $\dot{x}^2+\dot{y}^2+\dot{z}^2$ in the x-equation, those in $x\dot{y}-y\dot{x}$ in the y-equation, and the similar terms in the z-equation. But in such terms each of the coefficients is at the highest .1, so that no loss occurs when the product is divided by a number of the order .1.

Of the terms which have divisors with a factor m^2 the most important are the long period terms with arguments $2F-2l$, $2D+2l'-2l$, $2D+2l'-2F$ and the associated monthly terms with arguments which differ from these by $\pm\, l$.

For the first, the term with the lowest order characteristic is the monthly term with the argument $2F-l$, and this is therefore the first to be computed. Now in the x-equation (7. 6), the terms on the right have the explicit factors m^2, so that no loss arises from these. In the second degree terms on the left as before, each factor is at most 1 so that one place is lost from these. But as the term arises mainly from the terms with argument l, $2F$ in x, y and as the term with argument $2F$ can be taken to an extra place without difficulty we can avoid this loss. There is also a combination from $\dot{z}^2$ with arguments $l-F$, $-F$, but the former can also be taken to an extra place. Hence we can avoid all loss for the term with argument $2F-l$.

The next in order is the long period term with argument $2F-2l$. The additional part of the characteristic takes care of one decimal place, and the net loss is therefore one place.

In the succeeding monthly term there is no additional loss, because the only place where this might occur is in $\dot{x}^2 + \dot{y}^2$ and this involves the derivatives of the long period terms.

The second group of terms starts with the monthly term having the argument $2D + 2l' - l$. The only combination in the second order terms in which accuracy is lost is that with arguments $2D + 2l', l$. But the term with the former argument having a coefficient of the order .0001 can be taken to an additional place without trouble, so that loss of accuracy can be wholly avoided. In the succeeding long period term with argument $2D + 2l' - 2l$ we lose one place. With the monthly term $2D + 2l' - 3l$ no accuracy is lost.

The next term which needs consideration is that with argument $D + l' - l$ and having the characteristic $ee'a/a'$ so that it is of the fifth order. The preceding monthly term with argument $D + l'$ can be found without loss of accuracy because neither of the terms which might cause it is so great as .01. Hence proceeding as before we see that with this group only one place is lost.

Finally for the term $2D + 2l' - 2F$ which is the first of the group, the chief losses occur from the terms with argument $2D + 2l'$, $2F$ in x, y and $2D + 2l' - F$, F in z. As $2D + 2l' - F$ in z is a monthly term with a divisor m^2 there would be a loss if we do not take the term $2D + 2l'$ in x to an extra place, which we can. Hence z can be obtained without loss and the term with argument $2D + 2l' - 2F$ with a loss of one place. There is no loss with the succeeding short period terms.

Incidentally it may be noticed that there is no loss of accuracy in the coördinate x due to a preceding short period or long period term, and that the only loss of accuracy is in y when the division is of order m^2.

Hence in the whole theory, it is not now difficult to see that the maximum loss can be reduced to two places of decimals. As I started in T. M. M. with 11 places of decimals, it is probable that as far as the small divisors are concerned the theory is correct to 0″001 in longitude. This result is considerably better than I had hoped. With the equations actually used it was not easy to limit the apparent losses in the manner shown above. They may, however, have been avoided in the course of the arithmetic without the computer being aware of the fact. The verification to which reference was made earlier has proceeded sufficiently far to enable me to state that amongst the terms containing even multiples of D, F — about half of the whole number — there appears to be no error in the earlier work greater than 0″01 in longitude, on the assumption that the transformation from rectangular to polar coördinates has been correctly computed. Two of the terms containing odd multiples of D apparently have errors of ″.03, ″.04. But it is not certain whether these are apparent only, some further work being necessary to decide the matter.

YALE UNIVERSITY OBSERVATORY.

SUR LES AVANCES DU NOEUD ET DU PÉRIHÉLIE D'UNE PLANÈTE SOUS L'ACTION D'UNE ANNEAU CIRCULAIRE.

Par JEAN CHAZY, Paris.

1. Dans mon Livre *La théorie de la Relativité et la Mécanique céleste*, pour discuter les différentes explications possibles de l'avance observée du périhélie de Mercure par rapport à la théorie des grosses planètes, j'ai étudié[1] l'action séculaire d'un anneau circulaire et homogène ayant pour centre le corps central, sur les mouvements du nœud et du périhélie d'une planète: et l'on sait, selon le principe de la méthode de Gauss, que l'action d'une autre planète animée d'un mouvement circulaire[2] équivaut à l'action d'un anneau circulaire et homogène. J'ai appliqué la méthode de calcul classique fondée sur les éléments osculateurs et le système différentiel de Lagrange.

Dans un Mémoire posthume[3] qui rassemble ses recherches de Mécanique céleste, Fatou a étudié plus généralement les mouvements presque circulaires d'un point matériel dans un champ de gravitation présentant un axe de révolution et un plan de symétrie perpendiculaires, et dérivant d'ailleurs d'un potentiel newtonien, voisin le plus souvent du potentiel du corps central. Fatou applique des méthodes directes dans lesquelles les parties principales des avances du nœud et du périhélie peuvent se déduire de deux équations différentielles linéaires du second ordre, dont les coefficients sont presque constants.

Je veux ici, dans le but de comparer les deux sortes de résultats, et développant d'ailleurs trois Notes antérieures,[4] reproduire ou du moins résumer le calcul de Fatou.

2. Nous déterminons le champ considéré par le potentiel

$$F(r, z) = \mu/r + R(r, z)$$

par unité de masse: μ désigne le coefficient attractif du corps central O, r la distance d'un point matériel P au point O, z la cote du point P comptée sur l'axe de révolution Oz à partir du point O ou du plan de symétrie Π, et les deux fonctions F et R sont paires par rapport à la variable z. La fonction R est petite le plus souvent par rapport au terme μ/r. Les deux variables r et z conduisent à des formules plus simples que les coordonnées semi-polaires.

---

[1] *Loc. cit.*, t. 1 (1928), pp. 195-205.

[2] Tous les mouvements circulaires et tous les anneaux circulaires que nous considérons dans cet Article ont pour centre le corps central.

[3] *Acta Astronomica*, Série a, vol. 2 (1931), pp. 101-163.

[4] *Comptes rendus des séances de l'Académie des Sciences*, Paris, t. 203 (1936), pp. 706, 844 et 981.

Considérées comme fonctions des coordonnées cartésiennes correspondantes, F et R sont des fonctions harmoniques, et les équations de Laplace relatives à ces fonctions se transforment en les deux équations

$$\frac{\partial^2 F}{\partial r^2}+\frac{\partial^2 F}{\partial z^2}+\frac{2z}{r}\,\frac{\partial^2 F}{\partial r \partial z}+\frac{2}{r}\,\frac{\partial F}{\partial r}=0, \quad \frac{\partial^2 R}{\partial r^2}+\frac{\partial^2 R}{\partial z^2}+\frac{2z}{r}\,\frac{\partial^2 R}{\partial r \partial z}+\frac{2}{r}\,\frac{\partial R}{\partial r}=0, \tag{1}$$

équivalentes bien entendu l'une à l'autre; l'équation de Laplace relative à la différence

$$\frac{F-R}{\mu}=\frac{1}{r}$$

se réduit de même à l'identité à deux termes connue

$$\frac{\partial^2}{\partial r^2}\left(\frac{1}{r}\right)+\frac{2}{r}\,\frac{\partial}{\partial r}\left(\frac{1}{r}\right)=0.$$

Les deux dérivées $\partial^2 F/\partial z^2$, $\partial^2 R/\partial z^2$ ont évidemment la même valeur, et d'ailleurs Fatou a montré que leur valeur commune est négative dans les différentes applications.

Nous étudions des mouvements du point matériel P dont l'orbite est voisine d'une circonférence de centre O située dans le plan de symétrie Π, et qui sont d'ailleurs de sens direct.

En coordonnées semi-polaires, soient r_1, θ et z $(r=\sqrt{r_1^2+z^2})$, Fatou considère dans le mouvement du point matériel P l'intégrale des aires

$$r_1^2\theta'=C \qquad (C>0), \tag{2}$$

l'équation du mouvement relatif sur la projection du rayon vecteur sur le plan Π

$$r_1''-r_1\theta'^2=\frac{\partial F}{\partial r}\,\frac{r_1}{r},$$

ou

$$r_1''=\frac{C^2}{r_1^3}+\frac{\partial F}{\partial r}\,\frac{r_1}{r}, \tag{3}$$

et enfin l'équation de projection sur Oz

$$z''=\frac{\partial F}{\partial r}\,\frac{z}{r}+\frac{\partial F}{\partial z}. \tag{4}$$

En particulier soit un mouvement circulaire ayant lieu dans le plan de symétrie II, et où ρ et ω désignent le rayon et la vitesse angulaire : l'intégrale des aires (2) se réduit à

$$\rho^2\omega = C,$$

et montre que la vitesse angulaire ω est constante ; et l'équation (3) se réduit à l'équation

$$\frac{C^2}{\rho^3} + \frac{\partial F}{\partial \rho} = 0,$$

où $\partial F/\partial \rho$ désigne la valeur de la dérivée partielle $\partial F/\partial r$ pour $r = \rho$, $z = 0$. Et l'on tire de ces équations

$$\frac{1}{\rho}\frac{\partial F}{\partial \rho} = -\frac{C^2}{\rho^4} = -\omega^2. \tag{5}$$

3. En premier lieu, dans un mouvement qui tend vers le mouvement circulaire précédent, c'est-à-dire où la constante des aires C a exactement la même valeur que dans ce mouvement circulaire, et où le rayon vecteur r a des valeurs voisines du rayon ρ, calculons la durée de la révolution draconitique et l'avance du nœud. Dans le mouvement considéré, au second membre de l'équation (4), z est en facteur, puisque la dérivée $\partial F/\partial z$ s'annule avec z, d'une expression voisine de

$$\frac{1}{r}\frac{\partial F}{\partial r} + \frac{\partial^2 F}{\partial z^2},$$

et par suite de

$$\frac{1}{\rho}\frac{\partial F}{\partial \rho} - q,$$

ρ et ω étant définis par la constante C et par les équations (5), et $-q$ désignant la valeur commune des dérivées secondes $\partial^2 F/\partial z^2 = \partial^2 R/\partial z^2$ pour $r = \rho$, $z = 0$: q est positif puisque cette valeur commune est négative. Dès lors l'équation (4) peut s'écrire

$$z'' = -(\omega^2 + q)(1 + \epsilon)z, \tag{6}$$

ϵ désignant une quantité très petite.

Considérant z comme le seule fonction incomme, on sait que l'équation (6) s'intègre par deux quadratures ; Fatou étudie ces quadratures, mais, si nous nous bornons au terme principal de son calcul, nous pouvons dire que l'équation (6) est une équation linéaire en z dont les coefficients sont à peu près constants, et dont par suite les solutions sont voisines de fonctions périodiques de période $2\pi/\sqrt{\omega^2 + q}$.

En particulier la *durée de la révolution draconitique,* soit T_1, c'est-à-dire l'intervalle de temps compris entre deux instants consécutifs où la cote z s'annule en passant du négatif au positif, tend dans le mouvement considéré vers la même quantité

$$\lim T_1 = \frac{2\pi}{\sqrt{\omega^2 + q}}.$$

L'avance du nœud, soit $\delta\Omega$, c'est-à-dire la différence entre 2π et l'angle des rayons vecteurs correspondant aux deux instants qui précèdent, tend, puisque la vitesse angulaire tend elle-même vers ω, vers la valeur

$$\lim \delta\Omega = 2\pi\left(\frac{\omega}{\sqrt{\omega^2 + q}} - 1\right).$$

La fonction q étant positive en pratique, la valeur algébrique $\delta\Omega$ est négative: en fait, *dans le sens du mouvement, le nœud a, non une avance, mais un retard.*

4. En second lieu, concernant le terme principal de l'avance du périhélie, le calcul de Fatou peut se réduire à ceci. Soit un mouvement qui tend vers le mouvement circulaire considéré dans le plan II, mais ayant lieu lui-même dans ce plan. L'équation du mouvement relatif sur le rayon vecteur, soit

$$r'' = \frac{C^2}{r^3} + \frac{\partial F}{\partial r},$$

peut, comme précédemment l'équation de projection sur Oz, être considérée comme une équation différentielle linéaire dont les coefficients sont à peu près constants. En effet le second membre s'annule pour $r = \rho$, d'après la définition de ρ et parce que la constante C a la même valeur que dans le mouvement circulaire limite, et d'autre part il a pour dérivée pour $r = \rho$

$$-\frac{3C^2}{\rho^4} + \frac{\partial^2 F}{\partial \rho^2} = -3\omega^2 + \frac{\partial^2 F}{\partial \rho^2},$$

quantité que nous supposons négative: donc l'équation peut s'écrire

$$r'' = \left(-3\omega^2 + \frac{\partial^2 F}{\partial \rho^2} + \epsilon_1\right)(r - \rho),$$

ϵ_1 désignant une quantité très petite. Donc les solutions de la dernière équation sont voisines de fonctions périodiques dont la période est la quantité

$$\frac{2\pi}{\sqrt{3\omega^2 - \frac{\partial^2 F}{\partial \rho^2}}}.$$

En particulier la *durée de la révolution anomalistique*, soit T_2, c'est-à-dire l'intervalle de temps compris entre deux instants consécutifs où le rayon vecteur r est minimum, tend vers la même quantité. On peut d'ailleurs transformer cette quantité, et écrire, compte tenu des équations (5) et de l'équation de Laplace mise sous la forme (1),

$$3\omega^2 - \frac{\partial^2 F}{\partial\rho^2} = \omega^2 - \frac{2}{\rho}\frac{\partial F}{\partial\rho} - \frac{\partial^2 F}{\partial\rho^2} = \omega^2 - q,$$

$-q$ désignant comme ci-dessus la valeur de la dérivée $\partial^2 F/\partial z^2$ pour $r = \rho$, $z = 0$. La révolution a donc pour durée limite

$$\lim T_2 = \frac{2\pi}{\sqrt{\omega^2 - q}};$$

et l'*avance du périhélie*, soit $\delta\varpi$, c'est-à-dire la différence entre 2π et l'angle de deux rayons vecteurs minima consécutifs, tend vers la valeur

$$\lim \delta\varpi = 2\pi\left(\frac{\omega}{\sqrt{\omega^2 - q}} - 1\right).$$

q étant possitif, la valeur algébrique $\delta\varpi$ est positive aussi: *tandis que le nœud a, dans le sens du mouvement, un retard, le périhélie au contraire a effectivement une avance.*

5. Nous allons montrer encore que dans les deux cas les résultats obtenus peuvent être déduits de la méthode classique de détermination des petits mouvements d'un système matériel dont la position dépend d'un nombre fini de paramètres; et cette méthode donne même une extension du résultat obtenu dans le second cas.

En effet, dans le mouvement du point P, fixons encore la constante des aires C, éliminons l'angle polaire θ au moyen de l'intégrale des aires (2), et considérons dans le plan passant à chaque instant par l'axe Oz et par la position du point P, le mouvement relatif de ce point défini par le rayon vecteur r et la cote z, ou par les coordonnées cartésiennes r_1 et z.

La force relative dans ce mouvement est a priori la résultante de la force primitive, qui dérive de la fonction F, de la force centrifuge et de la force centrifuge composée correspondantes. Mais la force centrifuge est, au facteur -2 près, égale au produit vectoriel de la rotation instantanée du plan passant par Oz et par le point P, rotation dirigée suivant Oz, et de la vitesse relative du point P, vitesse située dans le même plan, qui est le plan du mouvement relatif que nous étudions. Donc la force centrifuge composée est perpendiculaire à ce plan, et par suite n'intervient pas dans le mouvement relatif considéré, en particulier y a un travail nul dans un déplacement virtuel quelconque.

La force centrifuge n'intervient que par sa projection sur le plan du mouvement: cette projection se déduit de l'accélération normale du point coïncidant, c'est la force perpendiculaire à Oz, et de valeur algebrique $\theta'^2 r_1$ dans le sens où l'on s'éloigne de Oz; son travail dans un déplacement virtuel du point P, de composantes δr_1, δz, a pour expression

$$\theta'^2 r_1 \delta r_1 = \frac{C^2}{r_1^3}\,\delta r_1 = \delta\left(-\frac{C^2}{2r_1^2}\right) = \delta\left[-\frac{C^2}{2(r^2-z^2)}\right].$$

La force relative dérive ainsi, en fonction des variables r et z, de la fonction

$$U = -\frac{C^2}{2(r^2-z^2)} + F(r, z).$$

On peut dire aussi que l'équation des forces vives du mouvement primitif, soit

$$r'^2_1 + r_1^2\theta'^2 + z'^2 = 2F + 2h,$$

h désignant la constante des forces vives, se transforme dans le mouvement relatif où l'on a

$$r_1^2\theta'^2 = \frac{C^2}{r_1^2} = \frac{C^2}{r^2-z^2};$$

en l'équation

$$r'^2_1 + z'^2 = -\frac{C^2}{r^2-z^2} + 2F + 2h,$$

de sorte que le travail des deux forces centrifuges dans un mouvement relatif réel quelconque, est égal nécessairement, à une constante additive près, à la quantité $-[C^2/2(r^2-z^2)]$.

La valeur $z=0$, une valeur commune ρ du rayon vecteur r et de la distance r_1 à l'axe Oz, et la vitesse ω donnée par l'équation $\rho^2\omega = C$, déterminent un mouvement circulaire du point P, si ces valeurs correspondent dans le mouvement relatif considéré à une position d'équilibre relatif. Or la dérivée

$$\frac{\partial U}{\partial z} = -\frac{C^2 z}{(r^2-z^2)^2} + \frac{\partial F}{\partial z}$$

est nulle identiquement avec z, puisque F est fonction paire de z: donc les valeurs ρ, ω et $z=0$ doivent annuler l'expression de la dérivée

$$\frac{\partial U}{\partial r} = \frac{C^2 r}{(r^2 - z^2)^2} + \frac{\partial F}{\partial r},$$

soit

$$\frac{C^2}{\rho^3} + \frac{\partial F}{\partial \rho} = \omega^2 \rho + \frac{\partial F}{\partial \rho} = 0,$$

équations équivalentes aux équations (5).

Pour étudier les mouvements du point P voisins du mouvement circulaire correspondant à cet équilibre relatif, nous sommes ramenés à étudier parmi les mouvements relatifs considérés les petits mouvements voisins de la position d'équilibre obtenue. Appliquons la méthode classique avec $r - \rho$ et z comme variables indépendantes.

La force vive relative est

$$r_1'^2 + z'^2 = \frac{(rr' - zz')^2}{r^2 - z^2} + z'^2;$$

et se réduit à $r'^2 + z'^2$ si l'on fait $r = \rho$, $z = 0$ dans les coefficients de r'^2, $r'z'$, z'^2. De même, dans la fonction de force réduite, le terme rectangle est nul, car la dérivée $\partial^2 U/\partial r \partial z$ est nulle à cause de la parité de F et de U; la dérivée $\partial^2 U/\partial z^2$ se réduit à la quantité négative

$$-\frac{C^2}{\rho^4} + \frac{\partial^2 F}{\partial z^2} = -\omega^2 - q,$$

selon la notation antérieure; enfin la dérivée $\partial^2 U/\partial r^2$ se réduit à la quantité

$$-\frac{3C^2}{\rho^4} + \frac{\partial^2 F}{\partial \rho^2} = -3\omega^2 + \frac{\partial^2 F}{\partial \rho^2} = -\omega^2 + q,$$

que nous transformons comme dans le calcul antérieur, et il suffit que cette dernière quantité soit négative pour que la position d'équilibre considérée soit stable. La fonction de forces réduite est ainsi, après multiplication par 2,

$$-(\omega^2 - q)(r - \rho)^2 - (\omega^2 + q)z^2.$$

La méthode classique donne alors les deux équations

$$(7) \qquad \begin{cases} r'' = -(\omega^2 - q)(r - \rho), \\ z'' = -(\omega^2 + q)z. \end{cases}$$

Nous nous trouvons donc dans le cas simple où, au lieu de former un système différentiel linéaire d'ordre 4, les équations des petits mouvements sont deux équations différentielles linéaires du second ordre déterminant séparément chacune une variable. Les périodes fondamentales sont les deux quantités

$$\frac{2\pi}{\sqrt{\omega^2 - q}} \quad \text{et} \quad \frac{2\pi}{\sqrt{\omega^2 + q}}$$

considérées précédemment. Donc, selon la proposition générale, dans les mouvements du point P voisins de mouvements circulaires, le rayon vecteur r et la cote z sont représentés par des développements suivant les puissances des valeurs initiales $r_0 - \rho$, z_0, r'_0, z'_0. Les termes linéaires de ces développements sont donnés par les deux équations linéaires (7), ce qui met en évidence que les durées des révolutions anomalistique et draconitique tendent respectivement, quand les valeurs initiales tendent vers zéro, vers les deux quantités $2\pi/\sqrt{\omega^2 - q}$ et $2\pi/\sqrt{\omega^2 + q}$, d'où se déduisent immédiatement les limites des avances du périhélie et du nœud. [Remarquons que ces limites de la durée de la révolution anomalistique et de l'avance du nœud sont valables dans un mouvement quelconque voisin d'un mouvement circulaire, c'est-à-dire sans supposer comme précédemment l'orbite située tout entiere dans le plan de symétrie du champ de gravitation].

Les quatre valeurs limites étant obtenues ainsi de façons très simples, je les formerai dans un autre Article en introduisant les éléments osculateurs du mouvement et le système différentiel de Lagrange.

PARIS.

DIE SYMMETRISCHEN PERIODISCHEN BAHNEN DES RESTRINGIERTEN DREIKÖRPERPROBLEMS IN DER NACHBARSCHAFT EINES KRITISCHEN KEPLERKREISES.

Von Ernst Hölder.

Beim restringierten Dreikörperproblem ergibt sich im folgenden die Mannigfaltigkeit der (zur Syzygienachse symmetrischen) periodischen Bahnen (K) in der Nachbarschaft eines („kritischen") Keplerkreises K^0 mit ganzzahligem Hillschen Periodenquotienten m als Sattelfläche

$$\text{(I)} \qquad \mu = \xi(\kappa - \tfrac{1}{2}a^{\frac{3}{2}}\xi^2 + \cdots)E(\kappa, \xi), \qquad E(0,0) = b \neq 0.$$

Diese wird von der Ebene verschwindenden Massenprozentsatzes μ im Punkte $\mu = \kappa = \xi = 0$ berührt und in den beiden Kurven geschnitten, die den Keplerkreisen $\xi = 0$ und den synodisch nach einem Umlauf geschlossenen Keplerellipsen entsprechen; κ ist der Zuwachs der Jacobischen Konstante von K gegenüber der von K^0, ξ im wesentlichen der $(m+1)$-te Fourierkoeffizient in der Kosinusentwicklung des Radiusvektor von K nach den Vielfachen der Elongation, ein Parameter, der sich für $\mu = 0$ näherungsweise auf die Exzentrizität reduziert; $a = (m/(m+1))^{2/3}$ ist der Radius von K^0.

Fixieren wir mit Wintner[1] den Massenprozentsatz $\mu > 0$, was der Strömgrenschen Gruppenfortsetzung entspricht, so ergeben sich auf der Lösungsfläche hyperbelartige Kurven, welche die von Poincaré[2] betonte analytische Fortsetzung der Lösungen einer natürlichen (isoenergetischen) Schar erster Sorte (über den kritischen Wert $\kappa = 0$ hinaus) in die Lösungen einer natürlichen Schar zweiter Sorte vermitteln—Scharen, die nach den Untersuchungen von Levi-Civita[3] bezw. Birkhoff[4] feststehen.

---

[1] A. Wintner, a) „Über eine Revision der Sortentheorie des restringierten Dreikörperproblems," *Ber. der Sächs. Akad. d. Wiss. zu Leipzig*, Bd. 72 (1930), S. 3-56; b) „Sortengenealogie, Hekubakomplex und Gruppenfortsetzung," *Mathematische Zeitschrift*, Bd. 34 (1931), S. 350-402. Auf diese Arbeiten stütze ich mich fortwährend, auch wegen ausführlicherer Literaturangaben sei auf sie verwiesen. Die von mir benutzte graphische Darstellung der Lösungsmannigfaltigkeit ist freilich eine andere als die auf der Massenschranke beruhende Wintnersche.

[2] H. Poincaré, „Les solutions périodiques et les planètes du type d'Hécube," *Bull. Astr.*, t. 19 (1902), S. 177-198, insbes. S. 185; die Figuren dort sind nur schematisch zu verstehen. Die zweite scheint mir auch nicht in allen Teilen die Tabelle von Brendel, „Theorie der Kleinen Planeten 1," *Abh. der Ges. d. Wiss. zu Göttingen*, Neue Folge 1, Nr. 2 (1898), S. 128, wiederzugeben.

Als Gegenstück werden auch die Verhältnisse in der Umgebung eines Keplerkreises K^0 mit rationalzahligem Periodenquotient $m = q/J$, $J > 0$, $(q, J) = 1$ wenigstens angedeutet, die sich ähnlich behandeln lassen. Dann gehört K^0 zwei verschiedenen Mannigfaltigkeiten von Lösungen an, welche die synodische Windungszahl 1 bezw. J haben, also zum ersten bezw. zweiten Poincaréschen Genre gehören und sich, wenn man wollte, in natürliche Scharen erster bezw. zweiter Sorte zusammenfassen ließen. Auch in jeder Ebene $\mu =$ const. existiert ein Doppelpunkt, durch den zwei verschiedene Lösungszweige hindurchgehen.

Die aus (I) approximativ folgende Beziehung

$$\xi = b^{-1}\mu\kappa^{-1},$$

im wesentlichen, hat sehr elegant Tisserand[5] hergeleitet — durch störungstheoretische Näherungsbetrachtungen, die sich aber wohl nicht so leicht zu einer strengen Methode für die Gewinnung des vollständigen analytischen Zusammenhangs ausgestalten ließen. Bei Tisserand ist zunächst nicht von periodischen Lösungen die Rede. Erst in einem kurzen Zusatz zur Darstellung seines Traité faßt er die ursprünglichen Entwicklungen von Radiusvektor und wahrer Länge als Fourrierreihen auf, deren Koeffizienten die in der störenden Masse linearen Approximationen der entsprechenden Koeffizienten einer periodischen Lösung sind. Diese hatte inzwischen, geleitet durch die Delaunaysche Mondtheorie, Hill[6] bemerkt und zugrundegelegt, als er im restringierten Dreikörperproblem Saturn—Titan—Hyperion, auf das übrigens auch Tisserand seine Theorie angewandt hatte, die Bewegung des Hyperion und die Titanmasse berechnete. Analytisch könnte die periodische Lösung mit derselben Methode der unendlich vielen Koeffizienten ermittelt werden, die Hill bei der Variationsbahn des Mondes entwickelt hat. Diese Methode hat neuerdings Wintner[7] mathematisch begründet und er hat auch die Anregung zu ihrer

---

[3] T. Levi-Civita, "Sopra alcuni criteri di instabilità," *Annali di Matematico* (3), vol. 5 (1905), pp. 221-308.

[4] G. D. Birkhoff, "The restricted problem of three bodies," *Rendiconti Palermo*, vol. 39 (1915), pp. 265-334.

[5] F. Tisserand, a) "Sur un cas remarquable du problème des perturbations," *Bull. Astr.*, t. 3 (1886), pp. 425-433; b) *Traité de la Mécanique Céleste*, t. 4 (1896), p. 113 ff., insbes. p. 116 f., sowie p. 491. Diese Untersuchung sowie die folgende von Hill hat Wintner, *loc. cit.* 1), a) hervorgehoben.

[6] G. W. Hill, "The motion of Hyperion and the mass of Titan" (1888), *Coll. Math. Works*, vol. 2 (1906), pp. 135-143.

[7] A. Wintner, a) "Zur Hillschen Theorie der Variation des Mondes," *Mathematische Zeitschrift*, Bd. 24 (1925), S. 259-265; b) "Über die Differentialgleichungen der

Anwendung auf das Problem der synodisch ganzzahligen Kommensurabilitäten gegeben, die von Martin[8] ausgeführt worden ist—allerdings nicht bis zu einem vollständigen Überblick über die Mannigfaltigkeit der Lösungen. Dazu muss man in der Rechnung noch einen Schritt weiter gehen und auch das Glied $\kappa\xi$ in (I) ermitteln. Ich tue das hier auf Grund der Lichtensteinschen[9] Theorie nichtlinearer Randwertprobleme, die mir[10] am nächsten liegt und deren ich vielleicht auch deshalb hier gedenken darf, weil sie ja eine grosse Analogie mit dem von Hill tatsächlich durchgeführten Verfahren hat: das Randwertproblem der symmetrischen Konjuktion und Opposition ist für symmetrische Bahnen mit dem der Periodizität gleichwertig, und die Berechnung der Störungen mittels mechanischer Quadraturen wird in der analytischen Theorie durch eine unter Zuhilfenahme der Greenschen Funktion gebildete Integrodifferentialgleichung ersetzt, die nach der Methode der sukzessiven Approximation gelöst wird.

1. Unter dem Einfluß des Newtonschen Gravitationsgesetzes sollen zwei Massenpunkte M_0 (Masse $1-\mu$, "Sonne") und M_1 (Masse μ, "Jupiter") umeinander Kreise beschreiben; der gemeinsame Radius der Relativbahnen sei gleich der Längeneinheit. Die Einheit der Zeit t sei so gewählt, daß die konstante Winkelgeschwindigkeit der Linie M_0M_1 gleich Eins ist.

In der Ebene der beiden Relativbahnen bewege sich ein Aufpunkt $P = P_\mu$ ("kleiner Planet"), welcher der Gravitationswirkung von M_0 und M_1 aus-

---

Himmelsmechanik," *Mathematische Annalen*, Bd. 96 (1926), S. 284-312; c) "Über die Existenz der Hillschen Mondbahn of maximum lunation und der Poincaréschen Schlingbahnen," *Mathematische Zeitschrift*, Bd. 28 (1928), S. 430-450; d) "Über die Konvergenzfragen der Mondtheorie," *Mathematische Zeitschrift*, Bd. 30 (1929), S. 211-227.

[8] M. Martin, "Upon the existence and non-existence of isoenergetic periodic perturbations of the undisturbed circular motions in the restricted problem of three bodies," *American Journal of Mathematics*, vol. 53 (1931), pp. 259-273.

[9] L. Lichtenstein, "Untersuchungen über die Gestalt der Himmelskörper, IV. Abh. Zur Maxwellschen Theorie der Saturnringe," *Mathematische Zeitschrift*, Bd. 17 (1923), S. 62-110; "V. Abh. Neue Beiträge zur Maxwellschen Theorie der Saturnringe," *Seeliger-Festschrift*, S. 200-225 (Berlin, 1925).

[10] Vgl. E. Hölder, a) "Mathematische Untersuchungen zur Himmelsmechanik," *Mathematische Zeitschrift*, Bd. 31 (1929), S. 197-256, insbes. S. 225-239. Eine Anwendung der Methode auf das Problem der synodisch ganzzahligen Kommensurabilitäten enthält meine Note b) "Die Verzweigungsgleichungen für die kritischen Kreise des restringierten Dreikörperproblems," *Ber. der Sächs. Akad. der Wiss.*, Bd. 83 (1931), S. 179-184. Aus dem dort gezeigten ausnahmslosen Nichtverschwinden des Koeffizienten von μ in der Verzweigungsgleichung folgerte ich damals mit Martin die Nichtexistenz einer von K^0 ausgehenden natürlichen Schar erster Sorte. Dieser Schluss muss aber ausdrücklich auf die von Hrn. Martin betrachteten Scharen eingeschränkt werden. Das vollständige Ergebnis wird erst hier gegeben und liefert eine positive Massenschranke auch für $\kappa \neq 0$, $\xi = 0$.

gesetzt ist, ohne diese Körper zu stören. Als Koordinaten von P nehmen wir die kartesischen Koordinaten x, y in dem rotierenden Achsenkreuz, dessen Ursprung in M_0 liegt und dessen x-Achse nach M_1 hin gerichtet ist.

Die Lagrangesche Funktion dieses restringierten Dreikörperproblems ist

$$(1)\quad \Lambda = \Lambda(x, y, \dot{x}, \dot{y}; \mu) = T + f + U,$$

wo

$$(2)\quad T = \tfrac{1}{2}(\dot{x}^2 + \dot{y}^2), \qquad f = x\dot{y} - y\dot{x},$$

$$(3)\quad U = \tfrac{1}{2}[(x-\mu)^2 + y^2] + (1-\mu)[x^2 + y^2]^{-\frac{1}{2}} + \mu[(x-1)^2 + y^2]^{-\frac{1}{2}}$$

und der Punkt die zeitliche Ableitung bezeichnet.

Da die Zeit t in Λ nicht explizit vorkommt, gilt das Jacobische Integral

$$(4)\qquad T - U = k.$$

2. Man gewinnt nun die zur Jacobischen Konstanten k gehörigen Bahnkurven der Gestalt nach als Extremalen des geometrischen Variationsproblems $\delta\int \Lambda^* d\tau = 0$ mit der Maupertuisschen Wirkungsfunktion

$$(5)\qquad \Lambda^* = [2(U + k)\cdot 2T^*]^{\frac{1}{2}} + f^*,$$

wobei T^*, f^* aus T, f entstehen, wenn man die Geschwindigkeitskomponenten $\dot{x}$, $\dot{y}$ durch die Ableitungen x', y' nach dem Parameter τ ersetzt. Dieser ist aus Homogenitätsgründen willkürlich und soll als der durch

$$(6)\qquad x = r\cos\tau, \qquad y = r\sin\tau$$

definierte Polarwinkel τ („Elongation") genommen werden, der also jetzt unabhängige Variable ist, während die gesuchte Funktion der Radiusvektor $r = r(\tau)$ ist. Es wird dann

$$(7)\qquad \Lambda^* = [2(U + k)(r'^2 + r^2)]^{\frac{1}{2}} + r^2.$$

Der zeitliche Ablauf $\tau = \tau(t)$ der Bewegung, der uns hier nicht weiter interessiert, bestimmt sich wegen (4) hinterher durch eine Quadratur aus

$$(8)\qquad \left(\frac{d\tau}{dt}\right)^2 = \frac{2(U + k)}{r'^2 + r^2}.$$

3. Wir untersuchen hier die Bahnen K, für die der Radiusvektor eine in $\tau \bmod 2\pi$ periodische Funktion $r(\tau)$ ist und die in der Nachbarschaft der zum Zweikörperproblem $\mu = 0$ gehörigen Keplerschen Kreisbahn

$$(9)\qquad K^0\colon\ x = a\cos\tau, \quad y = a\sin\tau$$

von P_0 liegen. Die Bahn K^0 hat die Jacobische Konstante, vgl. (51),

$$k^0 = -(2a)^{-1} - a^{\frac{1}{2}} \tag{10}$$

und die mittlere Bewegung n mit $n^2a^3 = 1$.

Der Charakter der gesuchten Lösungsmannigfaltigkeit (K) hängt wesentlich davon ab, ob für diese Ausgangslösung K^0 der Hillsche Periodenquotient

$$m = 1/(n-1) \tag{11}$$

eine ganze Zahl ist (singulärer, kritischer Fall) oder nicht (regulärer Fall). Die mittlere Bewegung ist jetzt $n = 1 + 1/m$.

Übrigens sind aus dynamischen Gründen die Werte

$$m = -1,\ 0,\ \infty,\ -\tfrac{1}{2} \tag{12}$$

auszuschließen; sie entsprechen bezw. den Grenzfällen, daß K^0 ins Unendliche rückt, in die Sonne hineinfällt, mit der siderisch direkten oder retrograden Jupiterbahn identisch ist.

Eine eingehendere Betrachtung bedarf nur der singuläre Fall; er kann nur für einen siderisch direkten Planeten eintreten. Denn die oberen und unteren retrograden Kreise gehören zu

$$-1 < m < -\tfrac{1}{2} \text{ bezw. } -\tfrac{1}{2} < m < 0, \tag{13}$$

können also niemals kritisch werden.

4. Für die zu untersuchenden Bahnen (K) sei der Radiusvektor

$$r = a(1+\zeta) \tag{14}$$

gesetzt; $\zeta = \zeta(\tau)$ muß die aus der Grundfunktion Λ^*, Formel (7), abgeleitete Lagrangesche Gleichung für ζ befriedigen. Daß K in einer Nachbarschaft von K^0 liegt, bedeutet, daß für K der Massenprozentsatz μ und die Differenz $\kappa = k - k^0$ der Jacobischen Konstanten absolut genommen kleine Werte besitzen und daß auch $\zeta(\tau)$ und seine Ableitung $\zeta'(\tau)$ dem Betrag nach unterhalb gewisser Schranken bleiben. Wir entwickeln am besten gleich Λ^*, wo auch $U = U(\tau, \zeta; \mu, \kappa)$ aufzufassen ist, nach Potenzen von μ, κ, ζ, ζ'. Dies wird für jeden einzelnen Wert von a gemacht.

Zum Zwecke der Entwicklung schreiben wir ausführlich

$$\begin{aligned} 2(U+k) = {} & a^2(1+\zeta)^2 - 2\mu a(1-\zeta)\cos\tau + \mu^2 - (2a^{\frac{1}{2}} + a^{-1}) + 2\kappa \\ & + 2(1-\mu)a^{-1}(1+\zeta)^{-1} \\ & + 2\mu[a^2(1+\zeta)^2 - 2a(1+\zeta)\cos\tau + 1]^{-\frac{1}{2}} \end{aligned} \tag{15}$$

und bekommen für die Lagrangesche Funktion den Ausdruck

$$(16)\qquad \Lambda^* = a[(V + 2W\mu + 2\kappa + \mu^2)(1 + 2\zeta + \zeta^2 + \zeta'^2)]^{\frac{1}{2}} + a^2(1 + \zeta)^2$$

mit

$$(17)\qquad V = a^2(1+\zeta)^2 + 2a^{-1}(1+\zeta)^{-1} - (2a^{\frac{1}{2}} + a^{-1}),$$

$$(18)\qquad W = [d^2 + 2a(a - \cos\tau)\zeta + a^2\zeta^2]^{-\frac{1}{2}} - a(1+\zeta)\cos\tau - a^{-1}(1+\zeta)^{-1};$$

hier ist

$$(19)\qquad d^2 = 1 - 2\,a\cos\tau + a^2$$

und, was sofort benutzt wird, (11) zufolge

$$(20)\qquad a^{-\frac{3}{2}} - a = am^{-1}, \qquad a^{-3} = (m+1)^2 m^{-2}.$$

In der Entwicklung von V und W brauchen wir, wie sich herausstellen wird, die Anfangsterme bis zum dritten bezw. ersten Grad in ζ:

$$(21)\qquad \begin{aligned} V_3 &= (a - a^{-\frac{1}{2}})^2 + 2(a^2 - a^{-1})\zeta + (a^2 + 2a^{-1})\zeta^2 - 2a^{-1}\zeta^3 \\ &= a^2m^{-2}\{1 + 2(1 - a^{-3})m^2\zeta + (1 + 2a^{-3})m^2\zeta^2 - 2a^{-3}m^2\zeta^3\} \\ &= a^2m^{-2}\{1 - 2(2m+1)\zeta + (3m^2 + 4m + 2)\zeta^2 - 2(m+1)^2\zeta^3\} \end{aligned}$$

und

$$(22)\qquad W_1 = d^{-1} - a^{-1} - a\cos\tau + (a^{-1} + (\cos\tau - a)ad^{-3} - a\cos\tau)\zeta.$$

Wir erhalten daher

$$(23)\qquad \begin{aligned} \Lambda^* &= a(V + 2W\mu + 2\kappa + \mu^2)^{\frac{1}{2}}(1 + \zeta + \tfrac{1}{2}\zeta'^2 - \tfrac{1}{2}\zeta\zeta'^2) + a^2(1+\zeta)^2 + \Lambda^{(1)} \\ &= aV_3^{\frac{1}{2}}(1 + \zeta + \tfrac{1}{2}\zeta'^2 - \tfrac{1}{2}\zeta\zeta'^2) + a^2(1+\zeta)^2 \\ &\quad + aV_1^{-\frac{1}{2}}W_1(1+\zeta)\mu + aV_2^{-\frac{1}{2}}(1 + \zeta + \tfrac{1}{2}\zeta'^2)\kappa - \tfrac{1}{2}aV_1^{-3/2}(1+\zeta)\kappa^2 + \Lambda^{(2)}. \end{aligned}$$

Dabei ist

$$(24)\qquad V_3^{\frac{1}{2}} = am^{-1}\{1 - (2m+1)\zeta - \tfrac{1}{2}(m^2 - 1)\zeta^2 \\ - \tfrac{1}{2}(m+1)(2m^2 + m + 1)\zeta^3 + \cdots\},$$

$$(25)\qquad V_2^{-\frac{1}{2}} = a^{-1}m\{1 + (2m+1)\zeta + \tfrac{1}{2}(9m^2 + 8m + 1)\zeta^2 + \cdots\}.$$

$V_1^{-\frac{1}{2}}$ ist der Abschnitt davon bis zu dem Glied mit ζ und

$$(26)\qquad aV_1^{-\frac{1}{2}}W_1 = m\{d^{-1} - a^{-1} - a\cos\tau + [a^{-1} + ad^{-3}(\cos\tau - a) - a\cos\tau \\ + (2m+1)(d^{-1} - a^{-1} - a\cos\tau)]\zeta + \cdots\};$$

$\Lambda^{(1)}$, $\Lambda^{(2)}$ und nachher $\Lambda^{(3)}$ enthalten außer einigen für die Lagrangesche Gleichung belanglosen, von ζ, ζ' unabhängigen Gliedern solche Glieder dritter

Ordnung in κ, μ, ζ, ζ', die den Faktor μ enthalten, ferner Glieder vierter und höherer Ordnung. Schließlich wird

$$
\begin{aligned}
(27)\qquad a^{-2}m\Delta^* = {}& \tfrac{1}{2}\zeta'^2 - \tfrac{1}{2}(m+1)^2\zeta^2 - (m+1)\zeta\zeta'^2 - m(m+1)^2\zeta^3 \\
& + a^{-2}m^2\{2(m+1)\zeta\kappa - a^{-2}m^2(3m+2)\zeta\kappa^2 \\
& + \tfrac{3}{2}(m+1)(3m+1)\zeta^2\kappa + \tfrac{1}{2}\zeta'^2\kappa \\
& + [2(m+1)d^{-1} - (2m+3)a\cos\tau \\
& + ad^{-3}\cos\tau - a^2d^{-3} - (2m+1)a^{-1}]\zeta\mu\} + \Lambda^{(3)} \\
= {}& \tfrac{1}{2}\zeta'^2 - \tfrac{1}{2}(m+1)^2\zeta^2 + \Sigma B_{ijkl}\kappa^i\mu^j\zeta^k\zeta'^l
\end{aligned}
$$

abgekürzt.

5. Die Lagrangesche Gleichung ist

$$
\begin{aligned}
(28)\qquad L(\zeta) \equiv {}& \zeta'' + (m+1)^2\zeta = \Sigma\{kB_{ijkl}\kappa^i\mu^j\zeta^{k-1}\zeta'^l - l\kappa^i\mu^j(B_{ijkl}\zeta^k\zeta'^{l-1})'\} \\
= {}& -\Sigma\{k(l-1)B_{ijkl}\kappa^i\mu^j\zeta^{k-1}\zeta'^l + l(l-1)B_{ijkl}\kappa^i\mu^j\zeta^k\zeta'^{l-2}\zeta'' \\
& + lB'_{ijkl}\kappa^i\mu^j\zeta^k\zeta'^{l-1}\} \\
= {}& -B_{0012}\zeta'^2 + (3B_{0030}\zeta^2 - 2B_{0012}\zeta\zeta'') + B_{1010}\kappa + B_{2010}\kappa^2 \\
& + (2B_{1020}\kappa\zeta - 2B_{1002}\kappa\zeta'') + B_{0110}\mu + \Psi^{(1)} = \Psi.
\end{aligned}
$$

Hier enthält $\Psi^{(1)}$ die Glieder zweiter Ordnung, soweit sie μ als Faktor haben, sowie die Glieder dritter und höherer Ordnung.

Die Ermittlung der in τ mod 2π periodischen Lösungen der Lagrangeschen Differentialgleichung (28) führt im regulären Falle $m \neq 1, \pm 2, \cdots$ mittels der Greenschen Funktion von $L(\zeta)$ zu einer nichtlinearen Integrodifferentialgleichung, deren Lösung $\zeta = \zeta(\tau; \kappa, \mu)$ sich für kleine Beträge von κ, μ durch sukzessive Approximationen ergibt und auch die Lagrangesche Gleichung (28) erfüllt und periodisch ist.

Im kritischen Fall $m = 1, \pm 2, \pm 3, \cdots$ hat $L(Z) = 0$ aber die periodischen Nullösungen $Z = \cos p\tau$, $\sin p\tau$ mit $p = m + 1$, und es muß mittels der verallgemeinerten Greenschen Funktion $G(\tau, \tau_*)$ zur nichtlinearen Integrodifferentialgleichung

$$
(29)\qquad \zeta(\tau) = \xi\cos p\tau - \eta\sin p\tau - \int_0^{2\pi} G(\tau,\tau_*)\Psi_*\,d\tau_*
$$

mit

$$
(30)\qquad \xi = \frac{1}{\pi}\int_0^{2\pi}\zeta\cos p\tau\,d\tau,\qquad \eta = \frac{1}{\pi}\int_0^{2\pi}\zeta\sin p\tau\,d\tau
$$

übergegangen werden.

6. Sind μ, κ und auch ξ, η beliebig vorgegeben in einem gewissen vierdimensionalen Interval Q um den Nullpunkt $\mu = \kappa = \xi = \eta = 0$, so ergibt sich durch sukzessive Approximationen genau eine Lösung $\zeta = \zeta(\tau; \mu, \kappa, \xi, \eta)$

von (29), die sich in eine Potenzreihe nach κ, μ, ξ, η entwickeln läßt. Damit die gewonnene Funktion ζ, die in τ mod 2π periodisch ist und die vorgegebenen Werte ξ, η als Parameter im Sinne von (30) hat, aber auch eine Lösung der ursprünglichen Lagrangeschen Differentialgleichung (28) ist, muß ζ noch die sog. Verzweigungsgleichungen

$$(31)\quad R \equiv (a/m)^2 \frac{1}{\pi} \int_0^{2\pi} \Psi \cos p\tau \, d\tau = 0, \quad S \equiv (a/m)^2 \frac{1}{\pi} \int_0^{2\pi} \Psi \sin p\tau \, d\tau = 0$$

befriedigen.

Schreiben wir in ζ die Glieder erster Ordnung, die aus der ersten Approximation entnommen werden können, aus, so kommt

$$(32)\quad \zeta = 2a^{-2} m^2 (m+1)^{-1} \kappa + \beta\mu + \xi \cos (m+1)\tau + \eta \sin (m+1)\tau + \cdots;$$

der Koeffizient β von μ ist für das folgende belanglos.

Zuerst berechnen wir

$$(33)\quad \begin{aligned} R = \{ & [12 B_{0030} + 4(m+1)^2 B_{0012}](m+1)^{-1} \\ & + (a/m)^2 [2 B_{1020} + 2(m+1)^2 B_{1002}]\}\kappa\xi + 2B^{(p)}\mu + R^{(1)} \\ = \{ & -12m(m+1) + (-4+1)(m+1)^2 + 3(m+1)(3m+1)\}\kappa\xi \\ & + 2B^{(p)}\mu + R^{(1)}, \end{aligned}$$

also

$$(34)\quad R = -6m(m+1)\kappa\xi + 2B^{(m+1)}\mu + R^{(1)},$$

wo

$$(35)\quad 2B^{(m+1)} = (a/m)^2 \frac{1}{\pi} \int_0^{2\pi} B_{0110} \cos (m+1)\tau \, d\tau$$

und $R^{(1)}$ sowie nachher $S^{(1)}$ Glieder zweiter Ordnung mit dem Faktor μ sowie Glieder dritter und höherer Ordnung enthält.

Die analoge Rechnung gibt

$$(36)\quad S = -6m(m+1)\kappa\eta + S^{(1)};$$

der Koeffizient von μ verschwindet hier offenbar aus Symmetriegründen.

7. Vor allem benötigen wir das Nichtverschwinden des Koeffizienten von μ, $2B^{(p)} = am^{-2} \cdot 2A^{(p)}$ in der Bezeichnung meiner früheren Arbeit b), in der ich den diesbezüglichen Nachweis allgemein erbracht habe.

Man führt dazu die Laplaceschen Koeffizienten A_s^p ein, die durch die Fourierentwicklung

$$(37)\quad (1 - 2\alpha \cos \tau + \alpha^2)^{-s} = A_s^0 + 2 \sum_{p=1}^{\infty} A_s^p \cos p\tau, \quad A_s^{-p} = A_s^p$$
$$(0 < \alpha < 1, \ s \text{ reell})$$

definiert sind. Sie können bekanntlich mit Hilfe der Gaußschen hypergeometrischen Reihe dargestellt werden

$$(38)\qquad A_s^p = A_s^{|p|} = \frac{\Gamma(s+|p|)}{\Gamma(s)\Gamma(|p|+1)}\,\alpha^{|p|}F(s, s+|p|, |p|+1, \alpha^2).$$

Für uns ist vor allem wichtig, daß für $0 < s$ sämtliche Glieder der Reihe positiv sind.

Wir setzen jetzt $s = \frac{1}{2}$ und unterscheiden:

1. Fall eines direkten unteren Planeten $\left(\substack{p=2,3,\dots\\ m=1,2,\dots}\right)$. Hier gilt mit $\alpha = a$

$$(39)\qquad 2B^{(p)} = 4pA_s^p - 2\alpha^2 A^p_{s+1} + \alpha(A^{p+1}_{s+1} + A^{p-1}_{s+1}).$$

Die A_s^p genügen nun bekanntlich den Identitäten

$$(40)\qquad \begin{aligned} 0 &= A_s^p - (1+\alpha^2)A^p_{s+1} + \alpha(A^{p+1}_{s+1} + A^{p-1}_{s+1}),\\ \frac{\alpha}{s}\frac{d}{d\alpha}A_s^p &= -A_s^p + (1-\alpha^2)A^p_{s+1}. \end{aligned}$$

Aus (39), $(40)_1$ ergibt sich durch Subtraktion

$$(41)\quad 2B^{(p)} = (4p-1)A^p_{1/2} + (1-\alpha^2)A^p_{3/2},\quad B^{(p)} = 2pA^p_{1/2} + \alpha\frac{d}{d\alpha}A^p_{1/2} > 0.$$

2. Fall eines direkten oberen Planeten $\left(\substack{p=-1,-2,\dots\\ m=-2,-3,\dots}\right)$. Dann ist mit $\alpha = 1/a$

$$(42)\qquad 2B^{(p)} = 4p\alpha A_s^p - 2\alpha A^p_{s+1} + \alpha^2(A^{p+1}_{s+1} + A^{p+1}_{s+1});$$

dies mit a multipliziert und davon $(40)_1$ subtrahiert, gibt

$$(43)\qquad \begin{aligned} 2aB^{(p)} &= (4p-1)A^p_{1/2} - (1-\alpha^2)A^p_{3/2},\\ aB^{(p)} &= (2p-1)A^p_{1/2} - \alpha\frac{d}{d\alpha}A^p_{1/2} < 0. \end{aligned}$$

(In der Arbeit b) hiess übrigens in diesem 2. Fall $aB^{(p)}$ einfach $B^{(p)}$).

8. Für $\mu = 0$ haben die Keplerschen Kreise verschwindende Parameter ξ, η. Die eigentlichen Keplerellipsen müssen die feste grosse Halbachse a, gegeben durch $(20)_2$ mit $m = 1, \pm 2, \cdots$, haben, um im synodischen Achsenkreuz x, y die Winkel-Periode 2π zu haben. Was die übrigen Elemente betrifft, so nehmen wir als Epoche $t = 0$ den Zeitmoment, wo die Bahn von P_0 die positive Syzygienachse kreuzt. Von dieser, als X_0-Achse siderisch fixiert, zählen wir (alles siderisch) die mittlere Länge $\lambda = nt$, die wahre Länge v und die Perihellänge ϖ; die mittlere Länge der Epoche ist $\epsilon = 0$. Die Exzentrizität e sowie $|\varpi|$ oder $|\varpi - \pi|$ seien klein vorausgesetzt. Sei noch die exzentrische Anomalie u, die Flächenkonstante $c = na^2(1 - e^2)^{\frac{1}{2}}$, so gelten die Beziehungen

$$(44)\quad \lambda - \varpi = u - e \sin u, \; r = a(1 - e \cos u), \; dv = cr^{-2}dt = (c/n)r^{-2}d\lambda.$$

Somit gelten für Radiusvektor und wahre Länge die Potenzreihen nach e:

$$(45)\quad r = a(1 - e \cos(\lambda - \varpi) + \cdots), \quad v = \lambda + 2e \sin(\lambda - \varpi) + \cdots.$$

Die Ellipsen, für die $|\varpi|$ bezw. $|\varpi - \pi|$ klein sind, bilden eine zusammenhängende stetige Schar, die man erhält, wenn man $|\varpi|$ klein und $-1 < e < +1$ nimmt.

Führen wir jetzt den von der Syzygienachse gezählten Polarwinkel ein:

$$(46)\quad \tau = v - t = \lambda(1 - n^{-1}) + 2e \sin(\lambda - \varpi) + \cdots,$$

$$(47)\quad p\tau = (m + 1)\tau = \lambda + 2pe \sin(\lambda - \varpi) + \cdots,$$

so ist umgekehrt

$$(48)\quad \lambda = p\tau - 2pe \sin(p\tau - \varpi) + \cdots,$$

$$(49)\quad \cos(\lambda - \varpi) = \cos(p\tau - \varpi) + 2pe \sin^2(p\tau - \varpi) + \cdots.$$

Nach $(45)_1$ wird jetzt für eine Keplerellipse die durch (14) definierte Grösse $\zeta = \zeta(\tau | e, \varpi)$ eine Potenzreihe in e, ϖ deren Koeffizienten (sogar schon mod $2\pi/p$) periodische Funktionen von τ sind. Unsere Parameter werden

$$(50)\quad \xi = -e + \cdots, \qquad \eta = -e\varpi + \cdots.$$

Es ist daher auch $\zeta = \zeta(\tau | \xi, \varpi)$ eine Potenzreihe in ξ, ϖ.

Für die betrachteten Keplerellipsen ist die Jacobische Konstante die Differenz von absoluter Energiekonstante h und Flächenkonstante c

$$(51)\quad k = h - c, \; h = h_0 = -1/(2a), \; c = c_0(1 - e^2)^{\frac{1}{2}}, \; c_0 = a^{\frac{1}{2}} = n^{-1/3}.$$

Es ist daher

$$(52)\quad \kappa = -(c - c_0) = -c_0[(1 - e^2)^{\frac{1}{2}} - 1] = \tfrac{1}{2}c_0\xi^2 + \cdots = \kappa(\xi, \varpi)$$

eine Potenzreihe in ξ, ϖ, die der Gleichung

$$(53)\quad P(\kappa, \xi, \varpi) \equiv \kappa - \tfrac{1}{2}c_0\xi^2 + \cdots = 0$$

genügt.

9. Bei der Diskussion der Verzweigungsgleichungen (31) muß ich mich—jedenfalls im Moment—auf die Bahnen beschränken, die zur Syzygienachse symmetrisch sind: $\zeta(-\tau) = \zeta(\tau)$. (Für $\mu = 0$ lassen wir in den Potenzreihen ζ, κ, P der vorigen Nr. das Argument $\varpi = 0$ einfach weg). Es ist dann von vornherein $\eta = 0$ zu setzen, und dies ist auch hinreichend dafür, daß die Lösung der Integrodifferentialgleichung (29) die Symmetrie bzgl. der Sygygienachse hat. Die zweite Verzweigungsgleichung $S = 0$ ist dann identisch in μ, κ, ξ erfüllt; es muß also S den Parameter η als Faktor enthalten.

Wir haben nur noch die Bedingung $R = 0$ für die Parameter μ, κ, ξ zu studieren. Da $B^{(p)} \neq 0$ ist, bestimmt sich der Massenprozentsatz als Potenzreihe in κ, ξ

$$(54) \qquad \mu = f(\kappa, \xi) = b\kappa\xi + \cdots, \qquad b = 3m(m+1)/B^{(m+1)} \neq 0.$$

Nun beachte man: die Potenzreihe $\zeta(\tau|\xi)$ in ξ zusammen mit der Potenzreihe (52) $\kappa = \kappa(\xi)$ und $\mu = 0$ in den Ausdruck R eingesetzt, gibt eine Potenzreihe in ξ, die identisch verschwindet. Auf der Lösungsfläche (54) im Raum der Koordinaten κ, ξ, μ liegen also jedenfalls die Punkte der parabelartigen Kurve (53) $P(\kappa, \xi) = 0$ in der Ebene $\mu = 0$, ferner, wie man ebenso sieht, die Punkte der Geraden $\xi = 0$, $\mu = 0$. Auf jeder der beiden im Ursprung irreduziblen algebroiden Kurven muß also sicher $f(\kappa, \xi) = 0$ sein. Daher gilt mit einer Potenzreihe $E(\kappa, \xi)$ die Teilbarkeit[11]

$$(55) \quad f(\kappa, \xi) = \xi P \cdot E = \xi(\kappa - \tfrac{1}{2}c_0\xi^2 + \cdots)E(\kappa, \xi), \quad E(0, 0) = b \neq 0.$$

Der Punkt $\kappa = \xi = 0$ ist ein Sattelpunkt der Lösungsfläche (54). Die Kurven $\mu = \text{const.}$ auf der Lösungsfläche haben in der Nähe von $\kappa = \xi = 0$ hyperbelartigen Charakter.[12]

Betrachtet man (etwa unter der Voraussetzung $m = 1, 2, \cdots$) die zum festen Wert $\kappa_1 < 0$ gehörige "natürliche" Lösungsschar Σ_1 (erster Sorte) sowie die zum festen Wert $\kappa_2 > 0$ gehörige natürliche Lösungsschar Σ_2 (zweiter Sorte mit $\xi < 0$)—natürlich nur innerhalb des Quaders Q —, so kann man wohl sagen, daß jeder Lösung von Σ_1 durch Gruppenfortsetzung auf der Lösungsfläche (also bei festem $\mu > 0$) eine Lösung von Σ_2 entspricht. Analog sind die Verhältnisse für positiven Exzentrizitätsparameter ξ, dem bei $\mu = 0$ die Perihellänge $\varpi = \pi$ entspricht. Nur sind jetzt beide $\kappa_1, \kappa_2 > 0$, sie können also auch gleich sein.

Auch die zum festen Wert $\kappa = 0$ gehörige natürliche Schar Σ_0 existiert. Es ist dann

$$(56) \qquad \mu = -\tfrac{1}{2}bc_0\xi^3 + \cdots,$$

also

$$(57) \qquad \xi = (-2b^{-1}c_0^{-1}\mu)^{1/3} + \cdots.$$

Diese vom kritischen Kreis ausgehende natürliche lineare Reihe Σ_0 von periodischen Lösungen entgeht der Betrachtungsweise von Martin, der $\mu = x^4$ schreibt und, *loc. cit.*, p. 267,

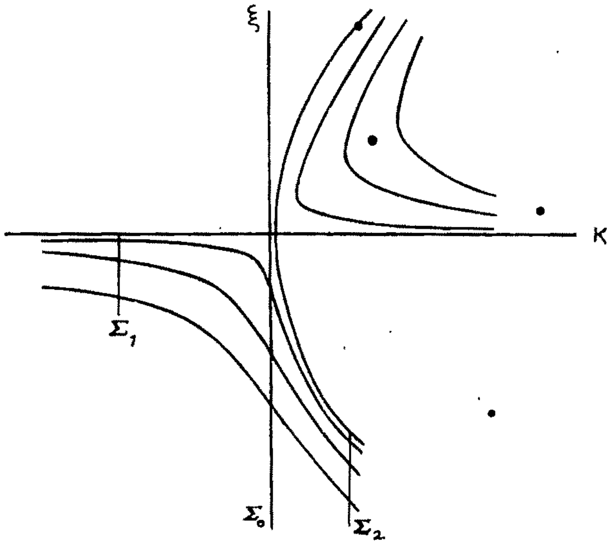
---

[11] Vgl. Osgood, *Lehrbuch der Funktionentheorie* II₁, Leipzig-Berlin 1924, p. 87, 2. Zusatz. Die entsprechende algebraische Tatsache gründet auf einem von meinem Vater bewiesenen Satz: O. Hölder, "Zum Invariantenbegriff," *Math.-naturw. Mitt. Württ.*, Bd. 1 (1884), S. 60-62.

[12] Vgl. Hadamard, *Cours d'analyse* I, p. 367 f., Paris, 1927.

$$\zeta = O(x^3) = O(\mu^{3/4}) \tag{58}$$

voraussetzt.



10. Auf der Lösungsmannigfaltigkeit liegt, wie noch gezeigt werden soll, die vom kritischen Kreis K^0 ausgehende isoperiodische Schar (der Lösungen von der synodischen zeitlichen Periode $T_0 = 2\pi m$) *in* der Ebene $\mu = 0$, sie besteht also nur aus den Keplerellipsen vom Periodenquotient m.[13]

Für die Lösungen der Integrodifferentialgleichung (29) ist die zeitliche Periode, durch (8) definiert,

$$T = T(\mu, \kappa, \xi) = \int_0^{2\pi} [2(U + k)]^{-\frac{1}{2}} a [\zeta'^2 + (1 + \zeta)^2]^{\frac{1}{2}} d\tau. \tag{59}$$

Setzen wir hier (54) zufolge $\mu = b\kappa\xi + \cdots$ ein, so ergibt sich (auf der Lösungsfläche)

$$T = \mathfrak{T}(\kappa, \xi). \tag{60}$$

Es geht nun

$$\frac{\partial \mathfrak{T}}{\partial \kappa} = \frac{\partial T}{\partial \kappa} + \frac{\partial T}{\partial \mu} \frac{\partial \mu}{\partial \kappa} \rightarrow \left(\frac{\partial T}{\partial \kappa}\right)_0 \text{ für } \kappa, \xi \rightarrow 0, \tag{61}$$

und es kann bei der an der Stelle $\mu = \kappa = \xi = 0$ zu nehmenden Ableitung rechts vor der Differentiation $\mu = \xi = 0$ gesetzt werden, d. h. man berechnet den Limes längs der von κ abhängigen Lösungsschar $\mu = \xi = 0$ von (29), die offenbar die Schar der zu $u=0$ gehörigen Keplerschen Kreise mit den Halbachsen $a(1+\gamma)$ sein muß. In dieser Schar ist

---

[13] Vgl. Wintner, *loc. cit.*, 1, a) S. 49.

$$(62)\quad \begin{aligned} T &= 2\pi(a^{-3/2}(1+\gamma)^{-3/2}-1)^{-1} = T_0(1-\tfrac{3}{2}a^{-3/2}(a^{-3/2}-1)^{-1}\cdot\gamma+\cdots)^{-1} \\ &= T_0(1+\tfrac{3}{2}(m+1)\gamma+\cdots) \end{aligned}$$

und

$$(63)\quad \begin{aligned} k &= -\tfrac{1}{2}a^{-1}(1+\gamma)^{-1}-a^{1/2}(1+\gamma)^{1/2} = k^0+\tfrac{1}{2}(a^{-1}-a^{1/2})\gamma+\cdots \\ &= k^0+\tfrac{1}{2}a^{1/2}m^{-1}\gamma+\cdots, \end{aligned}$$

also für $\kappa=0$

$$(64)\qquad \frac{\partial T}{\partial\kappa} = 3T_0(n+1)ma^{-1/2} = 6\pi m^3a^{-2} \neq 0.$$

Da sich in (61) nunmehr der Limes rechts von Null verschieden ergeben hat, geht durch $\kappa=\xi=0$ genau ein linearer regulärer Zweig $\kappa=\kappa(\xi)$ der Kurve $\mathfrak{T}(\kappa,\xi)=T_0$ hindurch. Dieser muß notwendig (52) $\kappa=\frac{1}{2}c_0\xi^2+\cdots$, $\mu=0$ sein.

11. Nur andeuten kann ich die Rechnungen, die auszuführen wären, um in der Nachbarschaft des Keplerkreises K^0 der mittleren Bewegung

$$(65)\qquad n=p/q,\quad p, q \text{ ganz},\ (p,q)=1,\quad p-q>0,$$

dieser $p-q=J$ mal syndosch durchlaufen, die (zur Syzygienachse symmetrischen) periodischen Bahnen zu übersehen, die sich (mindestens) nach dem synodischen Umlauf $2\pi J$ schließen. Es ist dann der Periodenquotient

$$(66)\qquad m=q/(p-q)=q/J \quad\text{und}\quad m+1=p/J, \qquad ((p,J)=1).$$

Multiplizieren wir die Differentialgleichung (28) mit J^2 und setzen

$$(67)\qquad \tau/J=\delta,$$

so kommt

$$(68)\qquad \frac{d^2\zeta}{d\delta^2}+p^2\zeta=\bar{\Psi}, \qquad (\bar{\Psi}=J^2\Psi).$$

Davon sind die im Intervall $0\leqq\delta\leqq 2\pi$ periodischen Lösungen zu suchen. Das zugehörige Intervall $0\leqq\tau\leqq 2\pi J$ heißt das primitive Intervall.

Was die Entwicklung der Lösung ζ der zu (68) gehörigen Integrodifferentialgleichung anlangt, so ändert sich, da ja die Differentialgleichung dieselbe geblieben ist, an dem konstanten Glied erster Ordnung gar nichts. Wir haben

$$(69)\quad \begin{aligned} \zeta &= 2a^{-2}m^2(m+1)^{-1}\kappa+\beta\mu+\xi\cos p\delta+\eta\sin p\delta+\cdots \\ &= 2a^{-2}m^2(m+1)^{-1}\kappa+\beta\mu+\xi\cos(m+1)\tau+\eta\sin(m+1)\tau+\cdots \end{aligned}$$

nach wie vor, wo jetzt

$$(70)\qquad \begin{aligned} \xi &= \pi^{-1}\int_0^{2\pi} \zeta \cos p\delta \, d\delta = (J\pi)^{-1}\int_0^{2\pi J} \zeta \cos\,(m+1)\tau d\tau, \\ \eta &= \pi^{-1}\int_0^{2\pi} \zeta \sin p\delta \, d\delta = (J\pi)^{-1}\int_0^{2\pi J} \zeta \sin\,(m+1)\tau d\tau \end{aligned}$$

und β ein wieder nicht in Betracht kommender Koeffizient ist.

Setzen wir von vornherein Symmetrie bezüglich der Syzygienachse, d. h. $\eta = 0$ voraus, so brauchen wir nur die eine Verzweigungsgleichung

$$(71)\quad \bar{R} = (a/m)^2\pi^{-1}\int_0^{2\pi} \Psi \cos p\delta \, d\delta = (a/m)^2\pi^{-1}J\int_0^{2\pi J} \Psi \cos\,(m+1)\tau d\tau = 0$$

zu betrachten. Darin berechnet sich der Koeffizient von $\kappa\xi$ in derselben Weise wie früher für $J = 1$, nur hat man zum Schluß das Integral

$$(72)\qquad \int_0^{2\pi J} \cos^2(m+1)\tau d\tau = J\int_0^{2\pi} \cos^2 p\delta \, d\delta = J\pi.$$

Ein Glied mit μ allein tritt in (71) für $J > 1$ nicht auf. Der Koeffizient $\bar{B}_{0110} = J^2 B_{0110}$ von μ in $\bar{\Psi}$ läßt sich nämlich in eine Kosinusreihe nach $\tau = J\delta$ entwickeln; ein Vielfaches dieses Winkels kann aber niemals gleich dem Winkel $p\delta$ sein, mit dessen Kosinus $\bar{\Psi}$ in (71) zu falten ist. Denn daß ein Vielfaches $NJ = p$ wird, ist bei der Teilerfremdheit von p und J nur für $J = 1$ möglich. Im ganzen bekommen wir für (71)

$$(73)\qquad \bar{R} \equiv J^2 \cdot 6m(m+1)\kappa\xi + \cdots = \xi(6pq\kappa + \cdots) = 0.$$

Dazu muß man noch beachten, daß $\bar{R} = 0$ gelten muß, sowie $\xi = 0$ ist. Denn für $\xi = 0$ hat die zu $0 \leqq \tau \leqq 2\pi J$ gehörige Integrodifferentialgleichung eine eindeutig bestimmte, in κ, μ reguläre analytische Lösungsschar. Diese muß man bekommen, wenn man von der zu $0 \leqq \tau \leqq 2\pi$ gehörigen Integrodifferentialgleichung ausgeht; deren Lösungen sind ja, da der reguläre Fall vorliegt, ohne weiteres Lösungen der Lagrangeschen Gleichung für die Bahnkurven und haben als Kosinusreihen in τ wieder wegen der Teilerfremdheit von p, J vermöge (70) den Parameter $\xi = 0$, genügen also der zu $0 \leqq \tau \leqq 2\pi J$ gehörigen Integrodifferentialgleichung mit $\xi = 0$.

Die Verzweigungsleichung (73) zeigt, daß die Mannigfaltigkeit der symmetrischen Lösungen zerfällt; auch bei festem μ erhalten wir zwei Lösungsgruppen, entsprechend Bahnen mit der synodischen Windungszahl 1 bezw. J.

Weitere Rechnungen bedürfte die Bestätigung der Vermutung, daß die Zeitspanne $JT_0 = 2\pi mJ = 2\pi q$ selbst als primitive synodische Periode von Bahnen mit $\mu > 0$ nicht vorkommt.

UNIVERSITÄT LEIPZIG.

SYMBOLIC DYNAMICS.

By MARSTON MORSE and GUSTAV A. HEDLUND.

1. Introduction. In both its qualitative and quantitative aspects, modern theoretical dynamics is largely concerned with recurrence and transitivity. The qualitative study of dynamics owes its inception to Hill and Poincaré. Its development in the past three decades has been dominated by the work of Birkhoff, to whom belongs the principal credit for the trend of modern research in dynamics. The quantitative (metrical) study of dynamical systems is of more recent origin, but with the establishment of the ergodic theorems of Birkhoff and von Neumann, its development has been rapid.

The methods used in the study of recurrence and transitivity frequently combine classical differential analysis with a more abstract symbolic analysis. This involves a characterization of the ordinary dynamical trajectory by an unending sequence of symbols termed a *symbolic trajectory* such that the properties of recurrence and transitivity of the dynamical trajectory are reflected in analogous properties of its symbolic trajectory. For general research of this character one may refer to papers by Hadamard, Birkhoff and the authors. Papers which are special in that they refer to spaces of constant curvature, but which are significant by virtue of the symbolisms involved have been written by Artin, Koebe, Martin, Myrberg and Nielsen. Martin and Robbins have contributed papers on purely symbolic aspects.

In the general theory as developed up to the present, the dividing line between symbolic and differential methods has not been sharply defined. It is a first aim of the present paper to separate the symbolic aspects from the differential aspects and to develop the symbolic theory on its own account. In this way many interesting and important dynamical questions can be given an algebraic formulation quite apart from the classical theories of differential equations.

A symbolic trajectory T is formed from symbols taken from a finite set of generating symbols subject to rules of admissibility which are dependent on the underlying Poincaré fundamental group. The trajectory T taken with one of its symbols is called a *symbolic element.* Symbolic elements are analogues of line elements on an ordinary trajectory. A metric is assigned to the space of symbolic elements and this space turns out to be perfect, compact, and totally disconnected. The necessary first theorems on symbolic recurrence and transitivity are then rapidly developed in a way suggestive of the differential treatment of Birkhoff.

We come to the more novel developments of the paper. Recall that the recurrency index $R(n)$ of T is a minimum number m such that each block of m successive symbols of T contains every block of n successive symbols of T. If $R(n)$ exists for each n, T is termed recurrent (cf. Morse [2], p. 94). In § 7 we are concerned with conditions on $R(n)$. A typical condition is that $R(n) > 2n$ for every non-periodic recurrent T. In § 8 we make a special study of the Morse recurrent trajectory (Morse [2], p. 95) obtaining $R(n)$ explicitly. This is the first and only known evaluation of $R(n)$ for a non-periodic recurrent trajectory. We give a definition of almost periodicity of symbolic trajectories somewhat similar to the ordinary definition of Besicovitch ([1], p. 77) and show that the Morse trajectory is not almost periodic. Most of the known recurrent trajectories are almost periodic. In § 9 we give general methods for deriving new recurrent trajectories from a given recurrent trajectory. For the Morse trajectory, $R(n)/n$ is shown to have an inferior limit of 5. 5 and a superior limit of 10. In § 10 we define a recurrent trajectory for which the inferior limit of $R(n)/n$ is finite while the superior limit is $+\infty$.

We continue with a study of transitivity. An admissible symbolic sequence of the form $a_1a_2a_3\cdots$ is termed a *ray*. We assign the Baire metric (cf. Hausdorff [1], p. 101) to the space of rays. Let $\theta(n)$ be the minimum length of a block which contains a copy of each admissible n-block. A ray R which contains a copy of each admissible block is termed transitive. If R is transitive, let $\phi(n)$ be the minimum length of an initial block of R which contains a copy of each admissible n-block. We generalize a theorem of Martin [1] and include a hitherto unpublished theorem of R. Oldenburger on the impossibility that $\phi(n) = \theta(n)$ for two successive values of n $(n > 1)$. A typical general theorem states that the set of transitive rays for which

$$\lim_{n\to\infty}\text{inf.}\,\frac{\phi(n)}{\theta(n)} = 1$$

is a residual set of the second category.

Symbolic dynamics as the authors conceive it forms one of the three divisions

(1) representation theory,
(2) symbolic dynamics,
(3) existence of space forms,

of the whole theory. The representation theory is concerned with the conditions on space forms under which trajectories admit a one-to-one symbolic representation in terms of which the recurrence or transitivity of the trajectory can be determined. These conditions will involve the Poincaré fundamental group of the space and differential conditions such as that of uniform instability (cf. Morse [4], p. 64). In (2) one is concerned with the existence of space forms satisfying the conditions discovered in (1). The questions involved are rather deep extensions of the Hilbert, Koebe theory of spaces of negative curvature (cf. Hilbert [1], and Koebe [1]). A simple typical theorem is that there exists no two-dimensional Riemannian manifold of the topological type of the torus satisfying the condition of uniform geodesic instability. The bearing of such studies on questions of topological and metric transitivity will be made clear in later papers.

2. Symbolic sequences. Having seen the significance of symbolic sequences we shall now develop their theory on an independent basis. We start with a finite set of μ distinct symbols termed *generating symbols.* We suppose that $\mu > 1$. If a_n represents a generating symbol, sequences of the form

$$\cdots a_{-1} a_0 a_1 \cdots, \tag{2.1}$$

$$a_r a_{r+1} a_{r+2} \cdots, \tag{2.2'}$$

$$\cdots a_{m-2} a_{m-1} a_m, \tag{2.2''}$$

$$a_{r+1} a_{r+2} \cdots a_{r+n}, \tag{2.3}$$

will be termed indexed *sequences* (written I-sequences). More particularly I-sequences of the forms (2.1), (2.2) and (2.3) will be termed *I-trajectories, I-rays,* and *I-blocks* respectively. It will be convenient to term the generating symbol represented by a_n the *value* of a_n. The I-blocks (2.3) will be said to have the *length* n and will frequently be termed n-blocks.

Suppose A and B are I-sequences of symbols a_n and b_m respectively. We shall regard A and B as identical and write $A \equiv B$ if and only if the ranges of the indices n and m are the same and a_n and b_n have the same value. Let r be an integer, positive, negative or zero. Suppose that the ranges of n and m are such that symbols a_n and b_m correspond in a one-to-one manner with a_n corresponding to b_{n+r}. If then a_n and b_{n+r} have the same value we shall say that A and B are *similar* and write $A \sim B$. We shall also write

$$B \equiv D_r A, \quad A \equiv D_{-r} B.$$

If $r = 0$, $D_r A \equiv A$, but in general $D_r A$ is different from A when $r \neq 0$.

The class of I-sequences similar respectively to an I-trajectory, I-ray, or I-block will be termed a *trajectory, ray,* or *block,* and will be represented by the sequence of values involved without indices. For example if 1 and 2 are generating symbols, the unending sequence

$$\cdots 121212 \cdots$$

is a trajectory but not an I-trajectory.

Corresponding to an I-sequence A we shall refer to *I-subblocks* or *I-subrays.* We shall thereby mean subsets of consecutive symbols of A 'indexed as' in A with successive indices differing by unity. Thus if A is of the form

$$a_1\, a_2\, a_3\, a_4,$$

a_2a_3 is an I-subblock, but a_1a_3 is not an I-subblock.

In classical dynamics an element based on a motion T is a point P of T together with the direction tangent to T at P. The symbolic analogue of such an element is defined as follows. Let (a) be an I-trajectory and a_r the r-th symbol of (a). The I-trajectory (a) taken with a_r will be termed an I-element based on (a) at a_r, and will be denoted by $E(r,a)$. We term r the *index* of $E(r,a)$ and a_r the *preferred* symbol. Let $E(s,b)$ be an I-element of index s based on an I-trajectory (b). We shall regard $E(r,a)$ and $E(s,b)$ as identical and write

$$E(r,a) \equiv E(s,b)$$

if and only if $(a) \equiv (b)$ and $r = s$. We shall say that $E(r,a)$ is similar to $E(s,b)$ and write $E(r,a) \sim E(s,b)$ if

$$D_{-r}(a) \equiv D_{-s}(b).$$

The class of I-elements similar to a given I-element $E(r,a)$ will be termed a *symbolic element E.* Two symbolic elements are regarded as identical if the corresponding classes of I-elements are identical. We say that E is based on (a) and represented by $E(r,a)$.

We shall give an illustrative example. Let (a) be an I-trajectory in which a_n equals 1 or 2 according as n is even or odd, and let (b) be an I-trajectory in which $b_{n+1} = a_n$. Evidently

$$E(0,a) \not\equiv E(0,b), \qquad E(0,a) \sim E(1,b).$$

More generally if n is even and m is odd,

$$E(n,a) \sim E(m,b)$$

and these similar elements serve to define a symbolic element E. There are just two different symbolic elements based on (a) or (b).

3. The space of symbolic elements. We shall assign a distance E_1E_2 to any two symbolic elements E_1, E_2. To that end we say that an I-block of the form

$$c_{r-m} \cdots c_r \cdots c_{r+m}$$

is *centered* at c_r. Let E_1 and E_2 be represented by indexed elements $E(r,a)$ and $E(s,b)$ respectively. If $E_1 \not\equiv E_2$, let m be the length of the longest I-blocks of (a) and (b) which are similar and centered at a_r and b_s respectively. We then set

$$E_1E_2 = E(r,a)E(s,b) = 1/m \qquad (m \neq 0). \tag{3.1}$$

In the special case where $m = 0$ we regard the distance (3.1) as infinite. If $E_1 \equiv E_2$, the distance (3.1) is taken as zero. It is clear that the distance E_1E_2 is independent of the particular I-elements $E(r,a)$ and $E(s,b)$ chosen to represent E_1 and E_2 respectively.

The distances so defined satisfy the usual metric axioms:

$$E_1E_2 = 0 \text{ if and only if } E_1 \equiv E_2, \tag{3.2}$$

$$E_1E_2 = E_2E_1, \tag{3.3}$$

$$E_1E_3 \leqq E_1E_2 + E_2E_3. \tag{3.4}$$

Relation (3.4) follows from the stronger relation

$$E_1E_3 \leqq \max(E_1E_2, E_2E_3). \tag{3.5}$$

In case any one of the distances involved in (3.5) is 0 or ∞, (3.5) holds trivially as is readily seen. Excepting these cases, let h, k, and m be the reciprocals of E_1E_3, E_1E_2 and E_2E_3 respectively. Consider the case $k \geqq m$. Evidently then $h \geqq m$ so that $h \geqq \min(k, m)$. The latter relation yields the condition

$$1/h \leqq \max(1/k, 1/m)$$

from which (3.5) follows. The case $m \geqq k$ leads to the same result.

Recall that a metric space M is compact if each infinite sequence of points of M contains at least one subsequence which converges to a point of M. With this understood we shall prove the following.

THEOREM 3.1. *The space M of all symbolic elements is compact.*

Let E_n be the n-th element in an arbitrary sequence of symbolic elements. Suppose E_n is based on the I-trajectory (c^n). Without loss of generality we can suppose that the preferred symbol in (a^n) has the index 0. We shall define a doubly infinite matrix of I-trajectories

$$T_{ij} \qquad (i, j = 1, 2, \cdots) \tag{3.6}$$

as follows. Since there is at most a finite number of generating symbols there exists an infinite subset of the symbols

$$a_0^n \qquad (n = 1, 2, \cdots)$$

which have the same "value" c_0. Let the I-trajectories T_{1j} be a subsequence of the I-trajectories (a^n) in which $a_0{}^n$ has the value c_0. Proceeding inductively we suppose that the first $k-1$ rows of the matrix (3.6) have been determined. The k-th row shall then be a subsequence T_{ki} of the $(k-1)$-st row such that the symbols of T_{ki} of index $k-1$ have a common value c_{k-1}, while the symbols of index $1-k$ have a common value c_{1-k}.

Let (c) be the I-trajectory

$$\cdots c_{-1}\, c_0\, c_1 \cdots.$$

Then (c) has an I-block of length $2k-1$, with center at c_0, in common with T_{kk}. Let $\mathcal{E}_k$ be the element of index 0 based on T_{kk}, and let $\mathcal{E}$ be the element of index 0 based on (c). Then

$$\mathcal{E}\mathcal{E}_k \leqq \frac{1}{2k-1},$$

so that the elements $\mathcal{E}_k$ converge to $\mathcal{E}$. Moreover the elements $\mathcal{E}_k$ are a subsequence of the given elements $\mathcal{E}_n$, and the proof of the theorem is complete.

A metric space is termed *totally disconnected* if its connected subsets all reduce to points. Cf. Hausdorff [1], p. 152. Recall also that if C is a connected set and d is a distance assumed by a pair of points of C, there exists a pair of points of C which assumes an arbitrary distance between 0 and d. In the case of the space M of symbolic elements, the distances take on only the discrete values $1/n$, so that the space M is totally disconnected.

Finally, the space M is perfect. We already know that M is compact. To prove that M is perfect it is necessary to show that in every neighborhood of an indexed element $E(0,a)$ there is an I-element $E(0,b)$ not similar to $E(0,a)$. To that end we take (b) so that it has a prescribed block A of (a) centered at a_0, in common with (a), but that all longer blocks of (a) with centers at a_0 differ from corresponding blocks of (b). The element $E(0,b)$ will not be similar to $E(0,a)$ but will have an arbitrarily small distance from $E(0,a)$ if the block A is sufficiently long. We accordingly have the following theorem.

THEOREM 3.2. *The space of all symbolic elements is compact, perfect and totally disconnected.*

By virtue of a well known theorem in point set theory (cf. Hausdorff [1], p. 160), Theorem 3.2 has the following corollary.

COROLLARY. *The space of all symbolic elements is the homeomorph of the Cantor, perfect, nowhere dense, linear set.*

4. The space of rays and the space of *I*-trajectories. Let R_1 and R_2 be rays with indexed representations of the form

$$a_r^{\bullet} a_{r+1} \cdots .$$

If $R_1 \not\equiv R_2$, the rays R_1 and R_2 have common initial blocks of finite maximum length m. With Baire (cf. Hausdorff, p. 101) we then set

$$R_1 R_2 = \frac{1}{m+1}.$$

If $R_1 \equiv R_2$, we set $R_1 R_2 = 0$. The distance so defined satisfies the usual metric axioms including a stronger triangle relation similar to (3.5). One readily sees that the space of rays is compact, perfect and totally disconnected.

This distance function in the space of rays is similar to that previously defined in the space of elements. We define a distance between I-trajectories which is based on an entirely different concept and is analogous to the Besicovitch metric of the theory of almost periodic functions (cf. Besicovitch [1], p. 73).

Let (a) and (b) be I-trajectories. Let $| a_i - b_i |$ be 0 if a_i and b_i have the same symbolic value, and 1 if this is not the case. We then set

$$[a, b] = \lim_{n \to \infty} \sup . \frac{1}{2n+1} \sum_{-n}^{n} | a_i - b_i | .$$

It is easily shown that this distance function satisfies the usual metric axioms with the exception that $[a, b] = 0$ does not imply $(a) \equiv (b)$. A simple example proving this last statement is obtained by letting all the symbols of (a) be 1, and all those of (b) be 1 with the exception of the symbol of index 0 which shall be 2.

If $(a') \sim (a)$ and $(b') \sim (b)$, it will not in general be true that $[a', b'] = [a, b]$, but corresponding to (a') there always exists a $(b') \sim (b)$ such that the equality does hold. Recalling the notation of § 2, let $D_r(a)$ denote the I-trajectory (a') in which $a'_{i+r} = a_i$. Thus $D_r(a) \sim (a)$. Furthermore, if $(a') \sim (a)$, there exists an r such that $D_r(a) \equiv (a')$. We shall establish the following lemma.

LEMMA 4.1. *If (a) and (b) are two I-trajectories and r is integral,*

$$[D_r(a), D_r(b)] = [a, b]. \tag{4.1}$$

Since $D_r(D_{-r}(a)) = (a)$, it is sufficient to prove (4.1) under the assumption that $r > 0$. Denoting $D_r(b)$ by (b') we have

$$\frac{1}{2n+1}\sum_{-n}^{n} | a'_i - b'_i | = \frac{1}{2n+1}\sum_{-n-r}^{n-r} | a_i - b_i |$$
$$= \frac{1}{2n+1}\sum_{-n}^{+n} | a_i - b_i | + \frac{1}{2n+1}\sum_{-n-r}^{-n-1} | a_i - b_i | - \frac{1}{2n+1}\sum_{n-r+1}^{n} | a_i - b_i |.$$

Since $0 \leqq | a_i - b_i | \leqq 1$, the last two terms approach 0 as n becomes infinite. Thus

$$\limsup_{n\to\infty} \frac{1}{2n+1}\sum_{-n}^{n} | a'_i - b'_i | = \limsup_{n\to\infty} \frac{1}{2n+1}\sum_{-n}^{n} | a_i - b_i |,$$

which is the stated result.

Let T be a trajectory and let (a) be an I-trajectory representing T. Given $\epsilon > 0$, T will be said to admit the *ϵ-translation* r if $[D_r(a), a] < \epsilon$. From (4. 1), this definition is independent of the I-trajectory chosen to represent T. Any trajectory admits the ϵ-translation 0, and this will be termed the *trivial* ϵ-translation.

As in the theory of almost periodic functions, a set of integers $\cdots < r_{-1} < r_0 < r_1 < r_2 < \cdots$ is termed *relatively dense* if the numbers $| r_i - r_{i-1} |$, $i = 0, \pm 1, \cdots$, form a bounded set. We define the trajectory T as *almost periodic* if given $\epsilon > 0$, T admits a relatively dense set of ϵ-translations. A periodic trajectory is evidently almost periodic. We shall presently define recurrent trajectories and shall use the notion of almost periodicity in the classification of such trajectories. Most of the known examples of recurrent trajectories turn out to be almost periodic, but the Morse recurrent trajectory of § 8 is not almost periodic, as we shall see.

5. Restricted sequences. In the applications to dynamical theory the sequences admitted are not formed freely from the generating symbols, but are subject to certain limitations. The simplest of these restrictions may be called the *restriction of inverses*. In this case the generating symbols form a set of the form

$$a_1, \cdots, a_p, a_1^{-1}, \cdots, a_p^{-1}, \qquad (p > 1)$$

in which a_i^{-1} is termed the inverse of a_i. In this case a sequence is inadmissible if a symbol and its successor are inverses. In other cases a circular order C of the generating symbols may appear and an integer r, with the restriction that p-blocks in the order C with $p > r$ are to be excluded. Another type of restriction prohibits successive blocks of certain definite forms.

In general problems restrictions of all of these types will appear. What these restrictions are in any problem depends upon the underlying group theory and topology. A detailed study of these restrictions will appear in a later paper. In the present paper we are fortunately able to abstract the essentials of these conditions as follows.

Certain blocks of the generating symbols will be distinguished and termed admissible. A ray or trajectory all of whose subblocks are admissible will be termed admissible as well as any element based on an admissible trajectory. The conditions of admissibility on blocks are as follows.

(α). *A block is admissible if and only if all its subblocks are admissible.*

(β). *Corresponding to any admissible block A there exist at least two generating symbols e, dependent at most upon the last two symbols of A and such that Ae is admissible. A similar condition holds replacing last by first and Ae by eA.*

(γ). *Corresponding to any two admissible blocks A and B there exists at least one generating symbol e dependent at most upon the last two symbols of A and the first two symbols of B and such that AeB is admissible.*

(δ). *There exist at least two generating symbols α and β such that all blocks formed from α and β alone are admissible.*

These conditions are satisfied when there are at least two generating symbols and their inverses and the only restriction is the restriction of inverses. They are also satisfied by the well known Nielsen symbolism (cf. Nielsen [2], pp. 211-217) for surfaces of genus $p > 1$. Various symbolisms developed by the authors satisfy these conditions, including one which replaces the Nielsen symbolism to advantage.

We state the following theorem.

THEOREM 5.1. *The set of admissible elements form a compact, perfect, totally disconnected subset Ω of the space M of all elements.*

That Ω is compact may be seen as follows. Let E_n be a sequence of admissible elements based respectively on trajectories t_n. Let E be a limit element of the sequence E_n. Suppose that E is based on the trajectory t. Each block in t must appear in some trajectory t_n and accordingly be admissible. Hence t and E are admissible, and Ω is compact.

To show that Ω is perfect let E be an arbitrary admissible element. Represent E by an I-element $E(0, i)$. Let A be a $(2n + 1)$-block of (a) centered at a_0. By virtue of admissibility condition (β), the block A can be extended to form an admissible I-trajectory of the form

$$(5.1) \qquad \cdots e_{-n-2}e_{-n-1}Ae_{n+1}e_{n+2}\cdots,$$

where the symbols e_i and e_{-i}, $i > n$, can be successively chosen in two different ways. Let E' be an element of index 0 based on (5.1). With the freedom of choice of the symbols e_i and e_{-i} it appears that the elements E' include elements different from E. But these elements also include elements arbitrarily near E since A is arbitrarily long.

The set Ω is accordingly perfect. It is totally disconnected since the space M is totally disconnected, and the proof of the theorem is complete.

A theorem similar to the above is true of the set of admissible rays, taking the metric of rays into account. Unless otherwise stated, the only trajectories and rays which will be used will be those which are admissible in the sense of this section.

6. Limit trajectories and minimal trajectories. Let t be a trajectory represented by an I-trajectory (a). If there exists a positive integer r such that

$$a_{n+r} = a_n \qquad (n = 0, \pm 1, \pm 2, \cdots),$$

t will be said to have the *period* r. A trajectory of period r is determined by each of its r-blocks. There are accordingly at most countably infinite many distinct periodic trajectories.

Let Σ be an arbitrary set of admissible trajectories. A trajectory t will be termed a *limit* trajectory of Σ if the elements based on trajectories of Σ include among their limit elements at least one element based on t. The set Σ can consist of a single trajectory, for example of the trajectory t given by

$$\cdots 2\,2\,2\,0\,2\,2\,0\,2\,0\,2\,2\,0\,2\,2\,2 \cdots. \tag{6.1}$$

The trajectory $\cdots 2\,2\,2 \cdots$ is a limit trajectory of the trajectory (6.1).

A periodic trajectory has no limit trajectory because it bears at most a finite number of elements. Let t be a non-periodic trajectory, and (a) an I-trajectory representing t. Any two I-elements with indices n and r respectively based on (a), with $n > r$, are dissimilar. For a similarity relation between these two elements would imply that (a) had the period $n - r$, contrary to hypothesis. The elements based on t with positive indices must then have at least one limit element, and we term such elements *positive limit elements* of t. *Negative limit elements* are similarly defined.

A trajectory can have more than one distinct limit trajectory. Letting a^n stand for a block $a \cdots a$ of length n we see that a trajectory of the form

$$\cdots \beta^4\alpha^3\beta^2\alpha\beta^2\alpha^3\beta^4 \cdots \tag{6.2}$$

has the trajectories $\cdots \alpha\alpha\alpha \cdots$ and $\cdots \beta\beta\beta \cdots$ as limit trajectories. More generally let the set of admissible blocks be enumerated in the form

$$A_0, A_1, \cdots.$$

If generating symbols e_n and b_n are successively chosen in accordance with block condition (γ) the trajectory

$$\cdots A_2 b_2 A_1 b_1 A_0 e_1 A_1 e_2 A_2 \cdots \tag{6.3}$$

will be admissible. It will have every admissible trajectory as a limit trajectory. The elements based on the trajectory (6.3) are everywhere dense in the set Ω of admissible elements. Such a trajectory is termed *transitive.*

In the remainder of this section we shall be concerned with a fundamental type of limit trajectories termed *minimal* trajectories. The dynamical counterpart of such trajectories was first introduced by Birkhoff [1]. Birkhoff also developed their general theory. In this connection the symbolic method was first used by Morse ([1], [2]) to establish the existence of non-periodic recurrent motions. Theorems 6.1 and 6.3 of this section and Theorem 7.2 of the following section are the abstract symbolic analogues of dynamical theorems of Birkhoff. The proofs presented here are independent of the theory of differential equations.

Let s be an admissible trajectory and n an arbitrary positive integer. The number of different n-blocks of s is finite. It will be denoted by $P_s(n)$ and termed the n-th *permutation number* of s. If t is a limit trajectory of s,

$$P_s(n) \geqq P_t(n). \tag{6.4}$$

A trajectory s such that the equality holds in (6.4) for every n and every limit trajectory t of s will be termed *minimal.*

We understand that periodic trajectories are minimal. They satisfy the above definition vacuously since they have no limit trajectories. We continue with a proof of the following theorem.

THEOREM 6.1. *Among the positive (or negative) limit trajectories of an admissible non-periodic trajectory T there is at least one admissible minimal trajectory.*

We shall give the proof for the case of positive limit trajectories. To that end we shall define certain symbols inductively as follows:

$H_0 =$ the set of positive limit trajectories of T.
$p_1 =$ the minimum of $P_t(1)$ for t in H_0.
$H_1 =$ the subset of H_0 for which $P_t(1) = p_1$.

. . . .

$p_n =$ the minimum of $P_t(n)$ for t in H_{n-1}.
$H_n =$ the subset of H_{n-1} for which $P_t(n) = p_n$.

The set H_0 is not empty since T is non-periodic. Hence H_n is not empty. Moreover

$$H_0 \supset H_1 \supset H_2 \cdots .$$

We continue with a proof of the following statement.

(α). *The set H_n is closed in the sense that it includes its limit trajectories.*

That H_0 is closed follows from the definition of limit trajectories. Proceeding inductively we shall assume that H_{n-1} is closed and prove that H_n is closed. In case the set H_n has no limit trajectories, H_n is certainly closed. Suppose then that H_n has a limit trajectory t. To show that t is in H_n, recall that $H_{n-1} \supset H_n$ and that t is accordingly a limit trajectory of H_{n-1}. Since H_{n-1} is assumed to be closed t is in H_{n-1}. By virtue of the definition of p_n,

$$P_t(n) \geqq p_n. \tag{6.5}$$

On the other hand, t is a limit trajectory of trajectories of H_n, and the n-th permutation numbers of trajectories of H_n equal p_n. Hence

$$P_t(n) \leqq p_n. \tag{6.6}$$

Hence the equality prevails in (6.5) and (6.6), and t is in H_n. Thus H_n is closed.

The sets H_n form a decreasing sequence of closed sets. It follows from the compactness of the space of admissible elements that the intersection H of the sets H_n is closed and non-void. Let s be a trajectory of H. If s is periodic, s is minimal and the proof is complete. If s is not periodic, let t be a limit trajectory of s. Since t and s are both in H_n for every n, t and s have the same n-th permutation number for every n. Hence s is minimal, and the proof is complete.

THEOREM 6. 2. *If s is a minimal trajectory and t a limit trajectory of s, then t is a minimal trajectory and s a limit trajectory of t.*

Every n-block of s is in t since s is minimal, so that s is a limit trajectory of t. To show that t is minimal, let r be one of its limit trajectories. Then r is also a limit trajectory of s. Since s is minimal, the n-th permutation numbers of t and r equal that of s, and so equal each other. Hence t is minimal, and the proof is complete.

The set of limit trajectories of a trajectory t will be termed the *derived set* of t. If s is a non-periodic minimal trajectory, each member of its derived set S has s as a limit trajectory, and hence has the same derived set. That this property is characteristic of non-periodic minimal trajectories is implied by the following theorem.

THEOREM 6.3. *A set S of trajectories which is the derived set of each of its members consists of non-periodic minimal trajectories.*

Since the derived set of a periodic trajectory is empty, no trajectory of the set S can be periodic.

If s is a trajectory of S and t is a limit trajectory of s,

$$P_s(n) \geqq P_t(n). \tag{6.7}$$

But t is in S and thus s is a limit trajectory of t, so that the equality sign must hold in (6.7). Hence s is minimal and the proof is complete.

The elements based on trajectories of S form a compact, perfect, totally disconnected set H. This is behind an earlier result of Morse that there exist discontinuous recurrent motions. Cf. Morse [2], p. 99. The set H is the homeomorph of the Cantor, perfect, nowhere dense, linear set. In particular H has the power $\aleph$ of the continuum. Since there is at most a countable infinity $\aleph_0$ of elements based on each trajectory of S we can conclude that the trajectories of S have a power a such that $\aleph_0 a = \aleph$. But $\aleph_0 a = a$ (cf. Hausdorff, pp. 30-31) so that $a = \aleph$. The set S thus has the power of the continuum.

We shall refer to the following definition. A set of trajectories which is the derived set of each of its members or which reduces to a single periodic trajectory is termed a *minimal set* of trajectories.

7. **Recurrent trajectories.** Let h be an admissible sequence, finite or infinite, and n a positive integer. A least integer m (if any exists) such that each m-block of h contains every n-block of h will be called the n-th *recurrency index* $R_h(n)$ of h. If h is a ray or a trajectory and $R_h(n)$ exists for each n, h will be termed *recurrent* and $R_h(n)$ the *recurrency function* of h.

We begin with the following theorem.

THEOREM 7.1. *If s is a recurrent trajectory and t is a limit trajectory of s, then t is recurrent and s is a limit trajectory of t. Moreover*

$$R_s(n) = R_t(n).$$

Each block of t appears in s since t is a limit trajectory of s. Let A be an arbitrary n-block of s. A copy of A is found in each block of s of length $R_s(n)$, and hence in each block of t of length $R_s(n)$. Hence $R_t(n)$ exists with

$$R_t(n) \leqq R_s(n). \tag{7.1}$$

Since each block A of s appears in t, s is a limit trajectory of t. The rôles of s and t can then be interchanged so that the equality prevails in (7.1). The proof of the theorem is complete.

The following theorem is the symbolic analogue of a theorem of Birkhoff. Cf. Birkhoff [2], p. 199.

THEOREM 7.2. *A necessary and sufficient condition that a trajectory s be recurrent is that it be minimal.*

We begin by assuming that s is minimal but not recurrent, and seek a contradiction. If s is not recurrent, there exists a block U of s with the following property. There exists a sequence A_n of blocks of s whose lengths become infinite with n and which do not contain U. Without loss of generality we can suppose that the length of A_n is odd. Let E_n then be the element based on s with preferred symbol at the center of A_n. Since the space of admissible elements is compact there exists a subsequence of the elements E_n which converges to an admissible element E. Let t be the trajectory on which E is based. It is clear that t does not contain the block U. Hence $t \neq s$. It follows that t is a limit trajectory of s. Since s is minimal, t and s must then contain the same blocks. From this contradiction we infer that s is recurrent.

It remains to show that s is minimal if recurrent. If s is periodic, s is minimal. Assume then that s is not periodic and let S be the derived set of s. It follows from Theorem 7.1 that S is the derived set of each of its members. We conclude from Theorem 6.3 that s is minimal, and the proof is complete.

It is natural to ask what limitations on recurrency indices $R(n)$ and permutation indices $P(n)$ are imposed by conditions of periodicity or recurrency. We give a partial answer in the following lemmas and theorems.

LEMMA 7.1. *If $P(n)$ is the number of different n-blocks in a trajectory t, $P(n+1) \geqq P(n)$.*

Let the different n-blocks in t be enumerated in the form

$$B_i \qquad [i = 1, 2, \cdots, P(n)], \tag{7.2}$$

and let e_i be the successor of B_i in some position in t. The $(n+1)$-blocks $B_i e_i$ are all different, and the lemma follows immediately.

If (a) is an indexed trajectory, a block B of (a) whose first and last symbols are a_r and a_m will be denoted by $[r, m]$. We shall term r the *first index* of B and m the *final index* of B.

THEOREM 7.3. *A trajectory t for which $P(m+1) = P(m)$ for some m has a period $\omega \leqq P(m)$.*

Let (a) be an indexed representation of t. There are $P(m)$ dissimilar m-blocks B_i in (a). The elements immediately following copies of B_i in (a) must always be the same. Otherwise there would be at least $P(m)+1$ dissimilar $(m+1)$-blocks. Similarly the elements immediately preceding copies of B_i must always be the same

Consider the blocks $B=[i+1, i+\mathrm{m}]$. Since there are just $P(m)$ dissimilar m-blocks, for some i there must exist a similarity relation of the form

$$(7.3) \qquad [i+1, i+m] \sim [i+1+\omega, i+m+\omega],$$

in which $0<\omega \leqq P(m)$. We shall show that (a) has a period ω. We begin with an inductive proof of the relations,

$$(7.4) \qquad a_{i+r}=a_{i+r+\omega} \qquad (r=1,2,\cdots).$$

That (7.4) holds for $r=1,2,\cdots,m$ follows from (7.3). Assuming that (7.4) holds for $r=1,2,\cdots,n \geqq n$, we shall prove that (7.4) holds for $r=n+1$. The m-blocks with final indices $i+n$ and $i+n+\omega$ respectively are similar by virtue of our inductive hypothesis. It then follows from the principle stated in the first paragraph of this proof that (7.4) holds for $r=n+1$.

It follows similarly that $a_{i+r}=a_{i+r+\omega}$ for $r \leqq 0$. We thereby make use of the fact that two similar m-blocks must be preceded by symbols with equal values. One starts with the similarity relation (7.3) and proceeds inductively. We conclude that (a) has the period ω, and the proof is complete.

COROLLARY. *For a non-periodic trajectory $P(m+1)>P(m)$ for each $m>0$.*

There exist trajectories for which $P(n+1)=P(n)+1$, so that the relation $P(n+1)>P(n)$ of the corollary is in a sense the best possible. An example of a recurrent trajectory for which $P(n+1)=P(n)-1$ for each $n>0$ will be given in a later paper.

LEMMA 7.2. *If t is a trajectory which involves μ generating symbols and has no period less than ω, then*

$$(7.5) \qquad P(n) \geqq n+\mu-1$$

for all values n for which $P(n)<\omega$.

Relation (7.5) is clearly true when $n=1$. From the preceding theorem $P(m+1)>P(m)$ for each m for which $P(m)<\omega$. Relation (7.5) follows inductively.

THEOREM 7.4. *If t is a non-periodic trajectory involving μ generating symbols,*

$$P(n) \geqq n + \mu - 1.$$

LEMMA 7.3. *If t is a recurrent trajectory with no period less than ω, then*

$$R(n) \geqq P(n) + n \geqq 2n + \mu - 1 \tag{7.6}$$

for each n such that $P(n) < \omega$.

Suppose that $P(n) < \omega$ and set $R(n) = m$. Since t is recurrent, each m-block B contains each of the $P(n)$ n-blocks of t. Hence

$$m \geqq P(n) + n - 1. \tag{7.7}$$

We shall exclude the equality in (7.7).

Suppose B has the form $[r+1, r+m]$ in an indexed representation (a) of t. Suppose $m = P(n) + n - 1$. Then no two n-blocks of B are similar. In particular the initial n-block A of B does not appear twice in B. But $R(n) = m$ so that A appears in the I-block $[r+2, r+m+1]$, and since A appears but once in B, A appears with final index $r+m+1$. Since A appears with final index $r+n$ as well, $a_{r+n} = a_{r+m+1}$ for arbitrary r. Hence t has a period

$$m - n + 1 = [P(n) + n - 1] - n + 1 = P(n) < \omega.$$

This is contrary to the hypothesis that t has no period less than ω. Hence $m > P(n) + n - 1$, and (7.6) follows with the aid of (7.5).

Lemma 7.3 and Theorem 7.4 yield the following theorem.

THEOREM 7.5. *If t is a non-periodic recurrent trajectory involving μ generating symbols,*

$$R(n) \geqq P(n) + n \geqq 2n + \mu - 1, \tag{7.8}$$

where $R(n)$ and $P(n)$ are respectively the recurrency and permutation indices of t.

It will be convenient to set $P(0) = 0$ for each trajectory t.

THEOREM 7.6. *If a trajectory t involves μ generating symbols and has a minimum period ω,*

$$R(n) \leqq n + \omega - 1 \qquad (n = 1, 2, \cdots). \tag{7.9}$$

Moreover there exists a largest integer m such that $P(m) < \omega$. For such an m,

$$(7.10)\qquad R(n) \geqq 2n + \mu - 1 \qquad (n = 1, 2, \cdots, m \leqq \omega - \mu),$$

$$(7.11)\qquad R(n) = n + \omega - 1 \qquad (n = m + 1, m + 2, \cdots).$$

Relation (7.9) is immediate. To establish the existence of a largest integer m such that $P(m) < \omega$, observe that each integer $n > 0$ for which $P(n) < \omega$ satisfies the relations

$$2n + \mu - 1 \leqq R(n) \leqq n + \omega - 1,$$

by virtue of (7.6) and (7.9). It follows that m exists and is at most $\omega - \mu$. Relation (7.10) follows from (7.6).

To establish (7.11) let n be any integer $> m$. Then $P(n) \geqq \omega$ and

$$(7.12)\qquad R(n) \geqq P(n) + n - 1 \geqq \omega + n - 1.$$

Relation (7.11) follows from (7.12) and (7.9).

Let (a) be an arbitrary I-trajectory. Let n be a positive integer, and let m be an integer, positive, negative or zero. For n and m fixed, let $H(m)$ be the minimum length of those blocks of (a) with initial index m which contain copies of all n-blocks of (a). If no such block with initial index m exists, let $H(m) = + \infty$. If, as m ranges over all integers, the set $H(m)$ is bounded, it is clear that the n-th recurrency index $R(n)$ of (a) exists and

$$(7.13)\qquad R(n) = \max. H(m).$$

Let r be any integer such that $H(r + 1)$ assumes the maximum of $H(m)$. The $R(n)$-block $[r + 1, r + R(n)]$ will then be termed a *minimax n-covering* in (a) or in the trajectory t represented by (a). The length of a minimax n-covering B in t is $R(n)$ and the final n-block in B does not otherwise appear in B. These two properties are characteristic provided $R(n)$ exists.

We are led to the following theorem.

THEOREM 7.7. *If t is a trajectory for which $R(n + 1)$ and $R(n)$ exist, $R(n + 1) > R(n)$.*

To prove the theorem let B be a minimax n-covering in t. Let A be the final n-block in B and e the symbol following A. The $(n + 1)$-block Ae does not appear in B since A has no second appearance in B. Hence $R(n + 1) > R(n)$, as stated.

It may happen that $R(n + 1) = R(n) + 1$ for infinitely many values of n as the recurrent trajectory of the following section shows.

8. The Morse recurrent trajectory. In this section we shall recall the definition of the non-periodic I-trajectory T first defined by Morse [2], p. 95, and shall give the explicit values of the corresponding recurrency function, denoting this function by $\rho(n)$.

Recall that the symbols α and β can be combined in any way to form admissible blocks. We here take α and β as 1 and 2 respectively. Set

$$\begin{array}{lll} & a_0 = 1, & b_0 = 2, \\ (8.0) & a_1 = a_0 b_0, & b_1 = b_0 a_0, \\ & \cdot\ \cdot\ \cdot\ \cdot & \cdot\ \cdot\ \cdot\ \cdot \\ & a_{n+1} = a_n b_n, & b_{n+1} = b_n a_n, \end{array}$$

understanding that the blocks $a_n b_n$ and $b_n a_n$ are to be expanded as blocks of 1's and 2's of length 2^{n+1}. The I-trajectory T shall have the representation

$$\cdots c_{-2} c_{-1} c_0 c_1 c_2 \cdots. \tag{8.1}$$

In T we set

$$c_0 c_1 \cdots c_{2^n-1} \equiv a_n. \tag{8.2}$$

Note that (8. 2) is consistent for all n, since a_m is an initial subblock of a_n for $m < n$. The symbols c_i are thereby defined for $i > 0$. To complete the definition of (8. 1) we set

$$c_{-r} = c_{r-1} \qquad (r > 0). \tag{8.3}$$

The subray

$$R = c_0 c_1 c_2 \cdots \tag{8.4}$$

of T thus starts as follows:

$$1221\ 2112\ 2112\ 1221\ 2112\ 1221\ 1221\ 2112 \cdots. \tag{8.5}$$

Corresponding to an arbitrary block A, the block which consists of the symbols of A in inverse order will be denoted by A^{-1}. We note that

$$\begin{array}{llll} (8.6') & a_n = a_n^{-1}, & b_n = b_n^{-1}, & (n \text{ even}) \\ (8.6'') & a_n = b_n^{-1}, & b_n = a_n^{-1}, & (n \text{ odd}). \end{array}$$

Relations (8. 6) are clearly true when $n = 0$, and their truth in general follows inductively from the relations

$$a^{-1}_{m+1} = (a_m b_m)^{-1} = b_m^{-1} a_m^{-1},$$
$$b^{-1}_{m+1} = (b_m a_m)^{-1} = a_m^{-1} b_m^{-1}.$$

From (8. 6) we infer that the 2^n-block of (8. 1) whose final index is -1 is $a_n^{-1} = a_n$ when n is even and $a_n^{-1} = b_n$ when n is odd. Thus the 2^{n+1}-block whose initial index is -2^n is $a_n a_n$ when n is even and $b_n a_n$ when n is odd. Each block of T appears in such a 2^{n+1}-block for n sufficiently large. Since

$$a_{n+3} = a_n b_n b_n a_n b_n a_n a_n b_n,$$

each block of T appears in the subray R of (8.4). The recurrency indices of T (if they exist) are thus completely determined by this subray.

Let A be an arbitrary block of 1's and 2's. The block obtained from A by replacing 1 by 2 and 2 by 1 will be denoted by A'. An easy inductive proof shows that

$$a'_n = b_n, \qquad b'_n = a_n. \tag{8.7}$$

The trajectories T' and T_n. Let T' be a trajectory with an indexed representation (e) in which $e_i = c_i$ for $i \geqq 0$, and for $i < 0$ equals 1 or 2 according as c_i is 2 or 1 respectively. The 2^n-block of T whose final index is -1 is a_n or b_n according as n is even or odd. The corresponding 2^n-block of T' is then $a'_n = b_n$ or $b'_n = a_n$ according as n is even or odd. We infer that each block of T' has a copy in the subray R of (8.4). The recurrency indices of T' (if they exist) are then completely determined by R. Since this is also true of T we have proved the following lemma.

LEMMA 8.1. *The recurrency indices of T and T', if they exist, are equal and are uniquely determined by the ray R in* (8.4).

We introduce a trajectory T_n, $n \geqq 0$, whose generating symbols are a_n and b_n, and which is obtained from T or T' according as n is even or odd, by replacing 1 by a_n and 2 by b_n. It follows from the relations (8.0), (8.3) and (8.6) that an expansion of the symbols a_n and b_n in T_n into blocks of 1's and 2's will yield T. Unless otherwise stated T_n shall be regarded as a trajectory in the symbols a_n and b_n unexpanded as blocks of 1's and 2's. For example $a_n b_n$ will be regarded as a 2-block of T_n, or, if we choose, as a 2^{n+1}-block of T.

We shall make use of the following properties of the ray R of (8.4).

(a). *There are no more than two successive* 1's *or* 2's *in R.*

(b). *Each 5-block of R contains the blocks* 12 *and* 21.

(c). *Each 5-block of R contains either the block* 11 *or else the block* 22.

To establish (a) we note that each 3-block of R will be found in one of the expanded 2-blocks of T_1; that is, in one of the blocks

$$a_1 b_1 = 1221, \qquad a_1 a_1 = 1212,$$
$$b_1 a_1 = 2112, \qquad b_1 b_1 = 2121,$$

and no more than two successive 1's or 2's appear in these blocks.

To verify (b) observe that a 5-block B which did not contain 12 would contain 21 in its first 3-block by virtue of (a). Its final 3-block would then be 111 contrary to (a). We infer that B contains 12, and similarly 21.

To prove (c) we infer from (a) that there can be no more than two consecutive a_1's or b_1's in T_1. In particular R cannot contain the block 12121 since R would then contain the block $a_1a_1a_1$ or $b_1b_1b_1$ according as 12121 possessed an even or odd first index. Similarly R contains no block 21212, and (c) follows directly.

Statements (a), (b) and (c) are true if R is replaced by T or T' since each block in T or T' appears in the subray R.

We continue with the following lemma.

LEMMA 8.2. *The recurrency index $\rho(n)$ of the trajectory T has the values*

$$\rho(1) = 3, \qquad \rho(2) = 9, \qquad \rho(3) = 11.$$

That $\rho(1) = 3$ follows from (a). We continue by showing that $\rho(2) \leqq 9$. To that end consider a 5-block B of T_1. The block B will contain the blocks

$$a_1b_1 = 1221, \qquad b_1a_1 = 2112, \tag{8.8}$$

by virtue of (b) and either

$$a_1a_1 = 1212 \quad \text{or} \quad b_1b_1 = 2121, \tag{8.9}$$

by virtue of (c). The only 2-blocks in T are 12, 21, 11 and 22. If B_0 denotes the block B expanded in terms of 1's and 2's, we see that each 2-block of T appears in B_0 *interior* to B_0. Thus the end symbols of B_0 can be deleted and the resulting block B'_0 contains all 2-blocks of T. But any 9-block of T will contain some block of the nature of B'_0 so that $\rho(2) \leqq 9$.

That $\rho(2)$ is not less than 9 follows upon considering the 9-block of T with initial index 10. This block has the form

$$12\,1221\,211,$$

and will contain no block 11, if the final 1 is removed. Thus $\rho(2) = 9$.

To show that $\rho(3) = 11$, recall that there are no more than two successive 1's or 2's in T so that the only 3-blocks are

$$112, \quad 122, \quad 121, \quad 211, \quad 221, \quad 212. \tag{8.10}$$

Consider the block B_0 of the preceding proof. It contains the blocks (8.8) and at least one of the blocks (8.9). It follows that B_0 contains each 3-block (8.10). But any 11-block of T will contain a B_0, so that $\rho(3) \leqq 11$.

That $\rho(3) \geqq 11$ follows upon noting that the 11-block of T with initial index 3 has the form

$$1\ 2112\ 2112\ 12,$$

and that removal of the final symbol from this block leaves no subblock of the form 212. Hence $\rho(3) = 11$, and the proof of the lemma is complete.

Let T'_1 be the trajectory obtained from T' by grouping the symbols of T' into 2-blocks with even first indices, regarding T'_1 as generated by the blocks a_1, b_1. It is readily seen that T'_1 is formed from the blocks a_1, b_1 exactly as T is formed from the symbols 1 and 2 respectively. We shall make use of this fact in proving the following lemma.

LEMMA 8. 3. *No two r-blocks of $T(T')$ for which $r \geqq 4$ are equal unless their first indices are equal mod 2.*

It is sufficient to prove the lemma for the case $r = 4$. To that end observe that each 4-block of $T(T')$ as it stands is in some 6-block A of $T(T')$ with even first index. Let A_1 be the 3-block of $T_1(T'_1)$ determined by A. Since $\rho(3) = 11$, each 3-block of $T_1(T'_1)$ appears in the 11-block of $T_1(T'_1)$ with first index 0. The lemma follows upon verifying its truth for the 22-block of T with first index 0. For each 4-block of T appears in this 22-block with a first index which is unchanged mod 2.

The following lemma is useful.

LEMMA 8. 4. *If $\rho(m+1)$ exists and $m \geqq 2$,*

$$\rho(2m) \leqq 2\rho(m+1) - 1.$$

We need only to show that each block B of T of length $2\rho(m+1) - 1$ contains a copy of each $2m$-block x of T. We distinguish two cases.

Case 1. *The first index of x is even.* Here x determines an m-block of T_1 and so appears in each expanded $\rho(m)$-block of T_1. Hence x appears in each $[2\rho(m) + 1]$-block of T. Thus $\rho(2m)$ exists and

$$2\rho(m) + 1 \geqq \rho(2m).$$

But

$$\rho(m+1) \geqq \rho(m) + 1,$$

in accordance with Theorem 7. 7. Hence

$$2\rho(m+1) - 1 \geqq 2\rho(m) + 1 \geqq \rho(2m).$$

This completes the proof in Case 1.

Case 2. *The first index of x is odd.* Let c' and c'' be respectively the elements of T which precede and follow the above block B. If B has an odd first index, $c'B$ has an even first index and a length $2\rho(m+1)$ and so contains a copy of each $(m+1)$-block of T_1. Hence $c'B$ must contain a copy y of x. But the index of y is odd in accordance with Lemma 8. 3, so that y does not include c'. Hence x appears in B. If B has an even first index, Bc'' contains a copy y of x. But the index of c'' is odd, while the final index of x is even. Hence y does not include c'' and lies in B.

The proof of the lemma is complete.

Since $\rho(3)$ exists it follows from the lemma that $\rho(4)$ exists, then $\rho(6)$, $\rho(10)$, etc. But if $\rho(n)$ exists, $\rho(m)$ exists for $m < n$. We conclude that $\rho(n)$ exists for all n.

We come to a principal theorem.

THEOREM 8. 1. *For $m \geqq 2$ the recurrency function $\rho(m)$ of T satisfies the relations*

$$\rho(2m) = 2\rho(m+1) - 1, \tag{8.11}$$

$$\rho(2m+1) = 2\rho(m+1). \tag{8.12}$$

Proof of (8. 11). It will be sufficient to show that the equality prevails in Lemma 8. 4. To that end let g be a minimax $(m+1)$-covering in T_1 as defined in § 7. The length of g is $\rho(m+1)$, and the final $(m+1)$-block of g appears only once in g. Let x be an arbitrary $2m$-block of T of odd first index, and let y be an $(m+1)$-block of T_1 which when expanded contains x. When $m \geqq 2$, x appears only as an interior block of the expanded y in accordance with Lemma 8. 3. Since a 2-block of T of even index is uniquely determined by either of its symbols we see that y is uniquely determined by x. It thus requires copies of all $(m+1)$-blocks of T_1 to cover all $2m$-blocks of T of odd index. In particular it will require a copy of the final $(m+1)$-block of g. As a block of T, g can then be shortened on the right by at most one symbol and, when expanded, still cover all $2m$-blocks of T. The equality accordingly holds in Lemma 8. 4, and the proof of (8. 11) is complete.

Proof of (8. 12). Recall that $\rho(2m+1)$ exceeds $\rho(2m)$. It follows then from (8. 11) that

$$\rho(2m+1) \geqq 2\rho(m+1). \tag{8.13}$$

It remains to show that the equality prevails in (8. 13). To that end let B be a $2\rho(m+1)$-block of T, and let c' and c'' be respectively the symbols preceding and following B. Let x be an arbitrary $(2m+1)$-block of T. We distinguish between several cases:

Case 1. The index of c' is even, that of x odd.
Case 2. The index of c' is even, that of x even.
Case 3. The index of c' is odd.

Case 1. The block x appears in some expanded $(m+1)$-block of T_1. Hence x has a copy y in $c'B$, for the first $2\rho(m+1)$-block of $c'B$ determines a $\rho(m+1)$-block of T_1. But the index of y is odd so that y does not include c' and thus lies in B.

Case 2. Here c'' has an odd index and the last $2\rho(m+1)$-block of Bc'' determines a $\rho(m+1)$-block of T_1. Hence x has a copy y in Bc''. But the final index of y is even so that y lies in B.

Case 3. In this case B determines a $\rho(m+1)$-block of T_1, hence contains a copy of each $(m+1)$-block of T_1, and accordingly a copy of x.

Thus a copy of x appears in B in all cases. The equality accordingly prevails in (8.13), and the proof of (8.12) is complete.

With the aid of the preceding theorem we can derive a general formula for $\rho(n)$ when $n \geqq 3$. To that end we first show that integers $n \geqq 3$ can be represented in the form

$$n = 2^r + p, \qquad (p = 2, 3, \cdots, 2^r + 1), \tag{8.14}$$

where $r \geqq 0$ and r and p are uniquely determined by n. The smallest and largest values of n given by (8.14) for a fixed r are respectively

$$2^r + 2, \qquad 2^{r+1} + 1,$$

and when r is replaced by $r+1$ the smallest value of n is

$$2^{r+1} + 2.$$

Thus as r ranges over the values $0, 1, 2, \cdots$ and p over the corresponding values listed in (8.14), n takes on each of the integers $3, 4, \cdots$ once and only once. We continue with the proof of the following theorem.

THEOREM 8.2. *For values of n of the form (8.14) the recurrency function $\rho(n)$ of T has the value*

$$\rho(n) = 10 \cdot 2^r + p - 1. \tag{8.15}$$

When $r = 0$ in (8.14) $p = 2$ and $n = 3$. We have seen that $\rho(3) = 11$, so that (8.15) holds when $n = 3$. When $r = 1$, we see that $p = 2$ and 3 while $n = 4$ and 5 respectively in (8.14). Relation (8.15) affirms that $\rho(4) = 21$ and $\rho(5) = 22$, and these values are correct as follows from (8.11) and (8.12) respectively upon setting $m = 2$. We proceed inductively assuming that (8.15) holds for a fixed r and the related values of p, and seek to prove that

$$\rho(\mu) = 10 \cdot 2^{r+1} + q - 1 \tag{8.16}$$

where

$$\mu = 2^{r+1} + q \qquad (q = 2, \cdots, 2^{r+1} + 1). \tag{8.17}$$

We distinguish between the cases where q is even and q is odd.

Case 1. *q even.* Here q has the form $2s$ and

$$\mu = 2(2^r + s).$$

Upon setting $m = 2^r + s$ in (8.11) we find that

$$\rho(\mu) = 2\rho(2^r + s + 1) - 1. \tag{8.18}$$

Upon referring to (8.17) we see that $s + 1$ ranges over the values

$$2, 3, \cdots, 2^r + 1. \tag{8.19}$$

When $n = 2^r + s + 1$, $p = s + 1$ in (8.14) and the range (8.19) is also the range prescribed in (8.14). We can accordingly use our inductive hypothesis to infer that

$$\rho(2^r + s + 1) = 10 \cdot 2^r + p - 1 = 10 \cdot 2^r + s. \tag{8.20}$$

Upon returning to (8.18) we conclude that

$$\rho(\mu) = 2(10 \cdot 2^r + s) - 1 = 10 \cdot 2^{r+1} + q - 1.$$

Relation (8.16) thus holds when q is even.

Case 2. *q odd.* Let q and μ be values of q and μ given by (8.17) with q odd. Corresponding to q and μ there exists a pair of even integers q_1, μ_1 again satisfying (8.17) with $\mu_1 = 2m$ and

$$q = q_1 + 1, \qquad \mu = 2m + 1.$$

Upon using the result established under Case 1,

$$\rho(\mu_1) = \rho(2m) = 10 \cdot 2^{r+1} + q_1 - 1.$$

But

$$\rho(2m + 1) = \rho(2m) + 1$$

by virtue of (8.11) and (8.12), so that in Case 2

$$\rho(\mu) = 10 \cdot 2^{r+1} + q_1 = 10 \cdot 2^{r+1} + q - 1.$$

Thus (8.16) holds in Case 2, and the proof of the theorem is complete.

This explicit determination of $\rho(n)$ is the first such determination for a non-periodic, recurrent trajectory.

The manner in which $\rho(n)$ varies with n is illustrated by the following table:

$n = 1,\ 2,\ 3,\ 4,\ 5,\ 6,\ 7,\ 8,\ 9,\ 10,\ 11,\ 12,\ 13,\ 14,\ 15,\ 16,\ 17,\ 18,$

$\rho(n) = 3,\ 9,\ 11,\ \underline{21,\ 22},\ \underline{41,\ 42,\ 43,\ 44},\ \underline{81,\ 82,\ 83,\ 84,\ 85,\ 86,\ 87,\ 88},\ 161,$

where the groups underlined are related in an obvious manner. For a given r in (8.14) the smallest and largest values of $\rho(n)$ are respectively

$$10 \cdot 2^r + 1, \qquad 11 \cdot 2^r.$$

We see that for a fixed r,

$$10 > \frac{10 \cdot 2^r + 1}{2^r + 2} \geqq \frac{\rho(n)}{n} = \frac{10 \cdot 2^r + p - 1}{2^r + p} \geqq \frac{11 \cdot 2^r}{2^{r+1} + 1}.$$

We are thus led to the following theorem.

Theorem 8.3. *If $\rho(n)$ is the recurrency function of T,*

$$\text{(8.21)} \qquad \lim_{n\to\infty}\text{inf.}\, \frac{\rho(n)}{n} = \frac{11}{2}, \qquad \lim_{n\to\infty}\text{sup.}\, \frac{\rho(n)}{n} = 10,$$

and 10 *and* 11/2 *are the limits respectively of $\rho(n)/n$ for sequences of values of n given by*

$$n = 2^r + 2, \qquad n = 2^{r+1} + 1, \qquad (r = 0, 1, \cdots).$$

If e is a prescribed positive constant and r is a sufficiently large positive integer, the ratios

$$\frac{\rho(n)}{n} = \frac{10 \cdot 2^r + p - 1}{2^r + p} \qquad (p = 2, 3, \cdots, 2^r + 1)$$

assume values within e of each value k between the limits (8.21). There accordingly exists a subsequence of values of n for which

$$\lim_{n\to\infty}. \frac{\rho(n)}{n} = k.$$

For all values of n, $\rho(n) < 10n$, while for each positive e and sufficiently large values of n, $2\rho(n) > (11 - e)n$.

That the trajectory T is not periodic follows from Theorem 7.6.

In § 4 a trajectory has been defined as almost periodic if it admits a relatively dense set of ϵ-translations for any positive ϵ. We show that a recurrent trajectory is not necessarily almost periodic by proving that the Morse recurrent trajectory admits only the trivial ϵ-translation if $\epsilon < 1/4$. To that end we first prove two lemmas.

LEMMA 8. 5. *If* $\cdots \delta_1\delta_0\delta_1 \cdots$ *is a sequence of the numbers* 0 *or* 1, *and* p *and* κ *are integers,* $\kappa > 0$, *then*

$$\limsup_{n\to\infty} \frac{1}{2n+1}\sum_{i=-n}^{n}\delta_i = \limsup_{n\to\infty}\frac{1}{2n+1}\sum_{i=-n}^{n+p}\delta_i = \limsup_{m\to\infty}\frac{1}{2\kappa m+1}\sum_{i=-\kappa m}^{\kappa m}\delta_i.$$

The first equality follows at once from the relationship

$$\lim_{n\to\infty}\left|\frac{1}{2n+1}\sum_{i=-n}^{n}\delta_i - \frac{1}{2n+1}\sum_{i=-n}^{n+p}\delta_i\right| \leqq \lim_{n\to\infty}\frac{p}{2n+1} = 0.$$

As to the second, let $n = \kappa m + r$, $0 \leqq r < \kappa$. Then

$$\frac{1}{2n+1}\sum_{i=-n}^{n}\delta_i = \frac{1}{2n+1}\sum_{i=-\kappa m}^{\kappa m}\delta_i + \frac{1}{2n+1}\sum_{i=-\kappa m-r}^{-\kappa m-1}\delta_i + \frac{1}{2n+1}\sum_{i=\kappa m+1}^{\kappa m+r}\delta_i,$$

where the last two sums are not present if $r = 0$. Each of these last two sums approaches 0 as n becomes infinite. It follows that

$$\limsup_{n\to\infty}\frac{1}{2n+1}\sum_{i=-n}^{n}\delta_i = \limsup_{n\to\infty}\frac{1}{2n+1}\sum_{i=-\kappa m}^{\kappa m}\delta_i$$
$$= \limsup_{n\to\infty}\frac{2\kappa m+1}{2n+1}\cdot\frac{1}{2\kappa m+1}\sum_{i=-\kappa m}^{\kappa m}\delta_i = \limsup_{m\to\infty}\frac{1}{2\kappa m+1}\sum_{i=-\kappa m}^{\kappa m}\delta_i.$$

The proof of the lemma is complete.

We recall that T_n is the trajectory with generating symbols a_n and b_n obtained from T or T', according as n is even or odd, by replacing 1 by a_n and 2 by b_n and that an expansion of the symbols a_n and b_n in T_n into blocks of 1's and 2's will yield T.

LEMMA 8. 6. *If* $r = 2^p q$, $p \geqq 0$, *and* (c) *and* (d) *are I-trajectories representing* T *and* T_p, *respectively,*

$$[D_r(c), (c)] = [D_q(d), (d)].$$

It follows from Lemma 4. 1 that the value of $[D_r(c), (c)]$ is independent of the particular indexed representation (c) of T which we use. We shall find it convenient to take (c) as the I-trajectory (8. 1). Then $D_r(c)$ is an I-trajectory (c') in which $c'_i = c_{i-r}$. Similarly we can suppose that in (d), d_i in expanded form is given by

$$d_i = c_{i2^p}\cdots c_{(i+1)2^p-1},$$

for (d) is then an indexed representation of T_p. Since $r = 2^p q$, $D_q(d)$ has the form (d') where $d'_i = d_{i-q} = c_{i2^p-r} \cdots c_{(i+1)2^p-1}$, in expanded form.

Since d_i and d'_i are either a_p or b_p, it follows from (8.7) that in their expanded form, d_i and d'_i are either identical, or one is obtained from the other by replacing 1 by 2 and 2 by 1. If we set

$$\{d'_i - d_i\} = \sum_{j=i2^p}^{(i+1)2^p-1} | c'_j - c_j |,$$

it follows that

$$\{d'_i - d_i\} = 2^p \mid d'_i - d_i \mid$$

With the aid of Lemma 8.5, we then see that

$$\begin{aligned}
[D_r(c), (c)] &= \limsup_{n\to\infty} \frac{1}{2n+1} \sum_{i=-n}^{n} | c'_i - c_i | \\
&= \limsup_{m\to\infty} \frac{1}{2\cdot 2^p m + 1} \sum_{i=-m2^p}^{m2^p} | c'_i - c_i | \\
&= \limsup_{m\to\infty} \frac{1}{2^{p+1} m + 1} \sum_{i=-m2^p}^{m2^p+2^p-1} | c'_i - c_i | \\
&= \limsup_{m\to\infty} \frac{1}{2^{p+1} m + 1} \sum_{i=-m}^{m} \{d'_i - d_i\} \\
&= \limsup_{m\to\infty} \frac{2^p}{2^{p+1} m + 1} \sum_{i=-m}^{m} | d'_i - d_i | \\
&= \limsup_{m\to\infty} \frac{1}{2m+1} \sum_{i=-m}^{m} | d'_i - d_i \quad = [D_r(d), (d)].
\end{aligned}$$

The proof of the lemma is complete.

We come to the proof of the desired theorem.

THEOREM 8.4. *The Morse recurrent trajectory admits only the trivial ϵ-translation* 0 *if* $\epsilon < 1/4$.

It is sufficient to show that

$$[D_r(c), (c)] \geqq \tfrac{1}{4}, \qquad r \neq 0, \tag{8.22}$$

where (c) is again the indexed representation (8.1) of the Morse recurrent trajectory.

Case 1. *r odd.* Let

$$\sigma_n = \sum_{i=n}^{n+3} | c'_i - c_i | = \sum_{i=n}^{n+3} | c_{i-r} - c_i |.$$

Since r is odd, the first indices of $c_{n-r}c_{n-r+1}c_{n-r+2}c_{n-r+3}$ and $c_n c_{n+1} c_{n+2} c_{n+3}$ are not congruent mod 2. It follows from Lemma 8.3 that these two 4-blocks cannot be similar, so that $\sigma_n \geqq 1$. Applying Lemma 8.5, we find that

$$\begin{aligned}[D_r(c), (c)] &= \limsup_{n\to\infty} \frac{1}{2n+1} \sum_{i=-n}^{n} | c'_i - c_i | \\ &= \limsup_{n\to\infty} \frac{1}{8n+1} \sum_{i=-4n}^{4n+3} | c'_i - c_i | \\ &= \limsup_{n\to\infty} \frac{1}{8n+1} \sum_{i=-n}^{n} \sigma_{4i}.\end{aligned}$$

Since $\sigma_n \geqq 1$, it follows that $\sum_{i=-n}^{n} \sigma_{4i} \geqq 2n+1$, and hence

$$[D_r(c), (c)] = \limsup_{n\to\infty} \frac{1}{8n+1} \sum_{i=-n}^{n} \sigma_{4i} \geqq \lim_{n\to\infty} \frac{2n+1}{8n+1} = \tfrac{1}{4}. \tag{8.23}$$

We have proved (8.22) when r is odd.

Let $(\bar{c})$ be the I-trajectory obtained from (c) by replacing c_n, $n < 0$, by 1 if c_n is 2, and by 2 if c_n is 1. Then $(\bar{c})$ is an indexed representation of the trajectory T'. Since Lemma 8.3 holds for T', as well as T, a proof similar to that given for (c) yields the relation

$$[D_r(\bar{c}), (\bar{c})] \geqq \tfrac{1}{4}, \qquad r \text{ odd}. \tag{8.24}$$

Case 2. *r even.* Since $r \neq 0$, $r = 2^p q$, where q is odd. From Lemma 8.6,

$$[D_r(c), (c)] = [D_q(d), (d)]. \tag{8.25}$$

The symbols of (d) are a_p and b_p, and if in (d) we set $a_p = 1$, $b_p = 2$, we obtain (c) or $(\bar{c})$ according as p is even or odd. Since q is odd, (8.23) or (8.24) applied to the right side of (8.25) yields

$$[D_q(d), (d)] \geqq \tfrac{1}{4}.$$

Combining this inequality with (8.25) we conclude that

$$[D_r(c), (c)] \geqq \tfrac{1}{4}, \qquad r \text{ even}.$$

The proof of the theorem is complete.

The following corollary is an immediate consequence of the preceding theorem.

COROLLARY. *The Morse recurrent trajectory is not almost periodic.*

Upon distinguishing recurrent trajectories according to whether they are almost periodic or not, we can state that recurrent trajectories of the type of

Morse are the only ones so far defined which are not almost periodic. In particular, the recurrent trajectories defined by Birkhoff ([4], p. 32) are almost periodic. The proof of this and the further development of the theory of almost periodic trajectories will be given in a later paper by the authors.

9. The derivation of recurrent trajectories. Having exhibited a non-periodic recurrent trajectory T, we shall describe four methods of deriving a recurrent trajectory from T, termed respectively derivation by projection, reduction, association and substitution. We also introduce a general method of derivation which includes each of the preceding. If the given trajectory consists exclusively of the free symbols α and β, these derived trajectories will all be admissible. In general they may or may not be admissible. If the given trajectory T is recurrent, the derived trajectories will be recurrent although usually differing in high degree from T. We suppose that T has an indexed representation (c) in terms of generating symbols $a_1, \cdots, a_\mu$.

Projection. The trajectory obtained from (c) by omitting a_k whenever it occurs in (c) will be said to be the k-th projection of (c).

Reduction. If the generating symbols consist of a finite set of integers, the trajectory obtained by reducing the elements c_j mod p will be said to be derived by reduction mod p.

Association. Let m be a positive integer at least 2, and let B_i be the m-block of (c) whose first index is i. There is at most a finite set of different blocks B_i. We introduce the I-trajectory

$$\cdots B_{-1}B_0B_1\cdots, \tag{B}$$

regarding the blocks B_i as the generating symbols.

Substitution. Let $A_1, \cdots, A_\mu$ be s-blocks of the generating symbols, $s > 0$. The trajectory (c^*) of a's obtained by replacing a_i in (c) by A_i and expanding will be said to be derived from (c) by block substitution.

Generalized substitution. Let $r_1, \cdots, r_m$ be a fixed set of integers, positive, negative, or zero. Let $\phi(x_1, \cdots, x_m)$ be a single-valued function (either numerical or symbolic) of its arguments x_i as each x_i ranges over the set $(a_1, \cdots, a_\mu)$ of generators of (c). We include the convention that a null set may be one of the symbolic values of ϕ. We introduce the symbols

$$E_i = \phi[c_{i+r_1}, c_{i+r_2}, \cdots, c_{i+r_m}],$$

and thereby define an I-trajectory (E). The number of different symbols E_i is finite. We add the convention that a null set E_i is to be dropped from the trajectory (E). Derivation by generalized substitution includes all of the preceding types of derivation. We shall illustrate this fact for the case of derivation by association. Derivation by association is obtained if we set

$$\phi(x_1, \cdots, x_m) = (x_1, \cdots, x_m),$$

and understand that

$$(r_1, \cdots, r_m) = (0, 1, \cdots, m-1).$$

The symbol E_i reduces to the block

$$B_i = c_i c_{i+1} \cdots c_{i+m-1},$$

as in the case of derivation by association.

We shall illustrate these processes further by deriving two recurrent trajectories from the trajectory T of the preceding section. We shall use the method of association letting $m = 2$. The generating symbols are the 2-blocks of T, namely 11, 12, 22, 21. We shall denote these blocks respectively by 1, 2, 3, 4 and write a few terms of the resulting trajectory. The symbols with indices $i \geqq 0$ written above the corresponding symbols of T begin as follows:

$$2342412341242342412$$
$$12212112211212212112.$$

Upon reducing the first of these trajectories mod 2 we obtain a trajectory which contains

$$0100010101000100010.$$

This trajectory is distinct from T since it has sequences of three equal symbols. These new trajectories are recurrent as we shall see.

A trajectory T' derived from a recurrent trajectory T by projection or reduction is clearly recurrent. The trajectory T' may however be periodic when T is non-periodic. The facts for the case of derivation by association are otherwise, as the following theorem shows.

THEOREM 9.1. *If (c') is a recurrent I-trajectory derived from a recurrent I-trajectory (c) by association in r-blocks, and $R'(n)$ and $R(n)$ are the recurrency functions of (c') and (c) respectively, then*

$$R'(n) = R(n + r - 1) - r + 1. \tag{9.1}$$

Moreover (c') *is periodic if and only if* (c) *is periodic.*

A block $[k + 1, k + n]$ of (c') is determined by the $(n + r - 1)$-block $[k + 1, k + n + r - 1]$ of (c) and conversely. We term these blocks *corresponding*. Each $(n + r - 1)$-block of (c) appears in each $R(n + r - 1)$-block H of (c). But H corresponds in the above sense to a block (c') of length

$$R(n + r - 1) - (r - 1), \tag{9.2}$$

while an $(n + r - 1)$-block of (c) corresponds to an n-block of (c'). Hence $R'(n)$ is at most the integer (9.2).

To show that $R'(n)$ is at least the integer (9.2), let H be a minimax $(n + r - 1)$-covering in (c), that is an $R(n + r - 1)$-block of (c) whose final $(n + r - 1)$-block appears just once in H. Let K be the block of (c') corresponding to H. The final n-block of K corresponds to the final $(n + r - 1)$-block of H and appears just once in K, so that $R'(n)$ is at least the integer (9.1), and the proof of (9.1) is complete.

If (c) is non-periodic, $R(n) > 2n$ in accordance with Theorem 7.5. It follows then from (9.1) that

$$R'(n) > 2n + r - 1.$$

The I-trajectory (c') is accordingly non-periodic, and the proof of the theorem is complete.

The preceding theorem leads to the following extension of Theorem 7.5.

THEOREM 9.2. *If* $P(r)$ *is the number of different* r*-blocks in a recurrent non-periodic* I*-trajectory* (c), *then for* $m \geqq r$,

$$R(m) \geqq 2m + P(r) - r. \tag{9.3}$$

Let (c') be derived from (c) by association in r-blocks. If $R'(n)$ is the recurrency function of (c'), it follows from Theorem 7.5 that

$$R'(n) \geqq 2n + P(r) - 1.$$

But from the preceding theorem we then see that

$$R(n + r - 1) - r + 1 = R'(n) \geqq 2n + P(r) - 1,$$

from which (9.3) follows upon setting $m = n + r - 1$.

Before coming to the next theorem we shall need the following definition.

A definition. Let there be given a finite set S of q-blocks $B_1, \cdots, B_p$ of a set of generating symbols. Let n be a positive integer less than q. A least integer $r(n) \leqq q$ (if such exists) such that each n-block appearing in a block of the set S appears in each $r(n)$-subblock of each of the blocks B_j will be called the n-th *recurrency index* of the set S. For example, let the blocks B_i be the blocks

$$B_1 = 1221, \quad B_2 = 2112, \quad B_3 = 1212, \quad B_4 = 2121. \tag{9.4}$$

These blocks have a recurrency index $r(1) = 3$. The recurrency index $r(2)$ does not exist.

The following theorem concerns the I-trajectory (c^*) derived from the I-trajectory (c) by substitution. The I-trajectory (c^*) is formed as described above by substituting the respective s-blocks A_i for the generating symbols a_i appearing in (c). These blocks A_i are then expanded to form the I-trajectory (c^*) of the symbols a_i. We shall apply the preceding definition of the recurrency index of a set of blocks to the set S of $2s$-blocks A_iA_j. Each m-block of (c^*) for which $m \leqq s+1$ lies in one of these $2s$-blocks of (c^*) so that in the case where the set S has a recurrency index $r(m) \leqq s+1$ it is clear that the recurrency index $R^*(m)$ of (c^*) exists and is at most $r(m)$. This is affirmed in part (a) of the theorem. Part (b) conditions $R^*(n)$ on the lower side instead of on the upper side. This theorem is useful in the following section.

THEOREM 9.3. *Let (c^*) be an I-trajectory derived from an I-trajectory (c) by substitution as described above.*

(a) *If the set of $2s$-blocks A_iA_j of (c^*) has an m-th recurrency index $r(m) \leqq s+1$, then*

$$R^*(m) \leqq r(m). \tag{9.5}$$

(b) *If some block A_i appears r times consecutively in (c^*) with $r > 1$, and if (c^*) does not have the period s, then*

$$R^*(s+1) > sr, \tag{9.6}$$

or else $R^*(s+1)$ fails to exist.

Statement (a) has already been proved. We therefore turn to the proof of (b).

It is assumed that the block $B = A_i \cdots A_i$, of length sr, appears in (c^*). Since the length of A_i is s, any $(s+1)$-block of B has equal terminal elements.

If $R^*(s+1)$ exists and $R^*(s+1) \leqq sr$, every $(s+1)$-block of (c^*) has a copy in B. Thus every $(s+1)$-block of (c^*) has equal terminal elements and (c^*) has the period s, contrary to hypothesis.

The proof of the theorem is complete.

The following theorem has a special interest in view of the mode of formation of the Morse recurrent trajectory. The trajectory H thereby affirmed to exist is not necessarily admissible. It will be admissible if the only symbols employed are α and β.

THEOREM 9.4. *Corresponding to an arbitrary recurrent ray*

$$R = a_0 a_1 a_2 \cdots,$$

there exists a recurrent trajectory H with R as subray and with the recurrency indices of R. The trajectory H is in general not uniquely determined by R.

Let B_r be the initial r-block of R, $r = 1, 2, \cdots$. The block B_r appears later in R immediately preceded by some r-block A_r. Of the blocks $A_1, A_2, \cdots$ there exists an infinite subset S_1 with last symbol in common. We take a_{-1} as this last symbol. Of the blocks of S_1 there exists an infinite subset S_2 with the second from the last symbol in common. We take a_{-2} as this common symbol. Proceeding in this way we define sets $S_3, S_4, \cdots$ and symbols $a_{-3}, a_{-4}, \cdots$ respectively. That the resulting trajectory H is recurrent and has the same recurrency indices as R follows from the fact that each block of H appears in R.

That H is not uniquely determined is shown by the trajectories T and T' of § 8 which have the recurrent ray R of (8.4) as a common subray.

10. Deferred recurrence. In this section we shall exhibit a recurrent trajectory E for which $R(n)$ exceeds a prescribed function $f(n)$ for infinitely many values of n but which is such that the limit inferior of $R(n)/n$ is finite. That a recurrent trajectory exists for which $R(n)$ exceeds a prescribed function $f(n)$ for all values of n has already been shown by Robbins ([1], Theorem 2). Our example shows the possibility of extreme variability in the indices $R(n)$. The trajectories E here defined form an extended class of which the trajectory T of § 8 is a member.

The generating symbols shall be the free symbols α and β. Let $r_0, r_1, \cdots$ be an arbitrary set of positive integers. We shall make use of powers of blocks, understanding that these powers are to be formally expanded. Set

$$\begin{aligned} A_0 &= \alpha, & B_0 &= \beta, \\ A_1 &= A_0^{r_0} B_0^{r_0}, & B_1 &= B_0^{r_0} A_0^{r_0}, \\ &\cdots\cdots & &\cdots\cdots \\ A_{n+1} &= A_n^{r_n} B_n^{r_n}, & B_{n+1} &= B_n^{r_n} A_n^{r_n}. \end{aligned}$$

Let s_n be the length of A_n and B_n. Observe that

$$s_{n+1} = 2s_n r_n. \tag{10.1}$$

Let (c) be an I-trajectory such that the initial s_n-block of the ray $c_0c_1c_2\cdots$ is A_n for each n. This definition is self consistent since A_{n-1} is an initial block of A_n. We further set

$$c_{-m-1} = c_m \qquad (m = 0, 1, \cdots). \tag{10.2}$$

Proceeding inductively we see that

$$\begin{array}{lll} A_n^{-1} = A_n, & B_n^{-1} = B_n, & (n \text{ even}) \\ A_n^{-1} = B_n, & B_n^{-1} = A_n, & (n \text{ odd}). \end{array}$$

Hence the s_n-block with final index -1 is A_n when n is even and B_n when n is odd. If the symbols of (c) are grouped into s_r-blocks with first indices $\equiv 0 \bmod s_r$, one obtains an I-trajectory E_r with generating symbols A_r and B_r.

If the integers r_i are all 1 and $\alpha = 1$, $\beta = 2$, (c) yields the recurrent trajectory T of § 8. We shall make use of the recurrency function $\rho(n)$ of T. The trajectory defined by (c) will be denoted by E. It is uniquely determined by the symbols α and β and the integers r_i.

THEOREM 10. 1. *The trajectory E is recurrent and non-periodic with a recurrency function $R^*(n)$ such that*

$$s_{n+1} \leqq R^*(s_n + 1) \leqq 2s_{n+3}. \tag{10.3}$$

Let B be an arbitrary $(s_n + 1)$-block of E. The block B appears in some expanded 2-block of E_n. But each such 2-block of E_n appears in both A_{n+3} and B_{n+3} as is easily seen. The latter two blocks are s_{n+3}-blocks of E and one of them at least is in each $2s_{n+3}$-block of E. Hence the right-hand condition in (10. 3) is valid. It follows that E is recurrent.

The left condition in (10. 3) follows from (b) in Theorem 9. 3. For the block A_n appears $2r_n$ consecutive times in E. Moreover E does not have the period s_n since E contains the block A_nB_n and B_n is different from A_n. From (9. 6) we then find that

$$R^*(s_n + 1) \geqq 2r_ns_n = s_{n+1},$$

and the proof of (10. 3) is complete. Moreover we see that

$$R^*(s_n + 1) \geqq 2s_n$$

and it follows from Theorem 7. 6 that E cannot be periodic.

COROLLARY. *If $f(n)$ is an arbitrary positive integral function, it is possible to choose the r_n's or any subset of the r_n's so that for corresponding values of n,*

$$R^*(s_n+1) > f(s_n+1). \tag{10.4}$$

Recall that $s_{n+1} = 2s_n r_n$. In order that (10.4) may hold we have merely to take r_n so large that

$$s_{n+1} > f(s_n+1)$$

and (10.4) will follow from (10.3).

We continue with the following theorem.

THEOREM 10.2. *If the sequence of integers r used in defining E contains arbitrarily long blocks of 1's,*

$$\liminf_{n\to\infty} \frac{R^*(n)}{n} \leqq \frac{11}{2}. \tag{10.5}$$

If $\rho(m)$ is the recurrency function of the trajectory T of § 8, there exists a sequence of integers m_i becoming infinite with i and such that

$$\liminf_{i\to\infty} \frac{\rho(m_i)}{m_i} = \frac{11}{2}. \tag{10.6}$$

We suppose integers m_i so chosen. The sequence of integers r_i contains arbitrarily long blocks of 1's. It follows that there exists a sequence of positive integers

$$h_1 < k_1 < h_2 < k_2 < \cdots, \tag{10.7}$$

such that for each i,

$$r_{h_i} = r_{h_i+1} = \cdots = r_{k_i} = 1. \tag{10.8}$$

Moreover these integers can be chosen so that

$$\mu_i = k_i - h_i > \rho(m_i).$$

We shall prove the following statement.

(α). *The m_i-th recurrency index of E_{h_i} is at most the recurrency index $\rho(m_i)$.*

The trajectory E_{h_i} is generated from the blocks A_{h_i}, B_{h_i}. These blocks combine to form the blocks A_{k_i}, B_{k_i} in exactly the way in which the symbols 1 and 2 combine to form the blocks a_{μ_i} and b_{μ_i} of the Morse trajectory T. In E_{h_i} the blocks

$$A_{k_i}B_{k_i}, \quad A_{k_i}A_{k_i}, \quad B_{k_i}A_{k_i}, \quad B_{k_i}B_{k_i}, \tag{10.9}$$

regarded as blocks of the symbols A_{h_i} and B_{h_i}, accordingly have an m_i-th recurrency index $\rho(m_i)$. For

$$\rho(m_i) < k_i - h_i < 2^{k_i - h_i} = 2^{\mu_i},$$

and 2^{μ_i} is the length of the blocks A_{k_i} and B_{k_i} regarded as blocks of the symbols A_{h_i}, B_{h_i}. It follows from (a) of Theorem 9.3 that E_{h_i} has an m_i-th recurrency index at most $\rho(m_i)$, and the proof of (α) is complete.

We now regard E as derived from E_{h_i} by expansion of the blocks A_{h_i} and B_{h_i} of E_{h_i} into blocks of α's and β's. Set

$$n_i = s_{h_i}(m_i - 1). \tag{10.10}$$

An arbitrary n_i-block of E is found in some expanded m_i-block of E_{h_i}, and hence in each expanded $\rho(m_i)$-block of E_{h_i}, and finally in each $[s_{h_i}\rho(m_i) + s_{h_i}]$-block of E. Hence

$$R(n_i) \leqq s_{h_i}[\rho(m_i) + 1]. \tag{10.11}$$

It follows from (10.10), (10.11) and (10.6) that

$$\lim_{i\to\infty}\text{inf.}\,\frac{R(n_i)}{n_i} \leqq \lim_{i\to\infty}\text{inf.}\,\frac{\rho(m_i)+1}{m_i - 1} = \frac{11}{2},$$

and the proof of the theorem is complete.

The preceding theorems and corollary combine to give use the following.

THEOREM 10.3. *If $f(n)$ is an arbitrary positive integral function of n, then for suitable choices of the integers r_i defining E, the recurrency function $R^*(n)$ of E exceeds $f(n)$ for infinitely many values of n while*

$$\lim_{n\to\infty}\text{inf.}\,\frac{R^*(n)}{n} \leqq \frac{11}{2}.$$

11. Transitive rays and their ergodic functions. We shall concern ourselves with rays of the form

$$c_0c_1c_2\cdots.$$

Such a ray will be termed transitive if it contains a copy of each admissible subblock. Transitive rays exist. If $A_1, A_2, \cdots$ is an enumeration of admissible blocks, the ray

$$X = A_1e_1A_2e_2\cdots$$

is admissible provided the symbols e_i are successively chosen in a suitable way. The ray X is clearly transitive. Let R be an arbitrary ray and let n be a positive integer. Suppose an initial r-block B of R contains each admissible n-block while no initial subblock of B has this property. The integer r will then be termed the n-th *transitivity index* $\phi(n)$ of R. If R is transitive, the n-th transitivity index of R exists for each value of n and will be termed the *ergodic function* $\phi(n)$ of R.

A block $H(n)$ of minimum length containing each admissible n-block will be termed a *minimum n-covering*. The length $\theta(n)$ of $H(n)$ will be termed the n-th *covering index*. In the case where the generating symbols are all free, Martin [1] has shown that for each n there exists a minimum n-covering in which no n-block is repeated. We shall presently extend Martin's result to the case where successive inverses are prohibited. We are concerned here with the existence of transitive rays whose ergodic function is in some sense relatively small. Noting that each ergodic function $\phi(n)$ satisfies the condition

$$\phi(n) \geqq \theta(n),$$

we are led to the question, does there exist a transitive ray such that $\phi(n) = \theta(n)$ for each n? The answer is no, in general, as we shall see.

We begin with the case where there are just two generating symbols, namely the free symbols α and β. In this case it follows from Martin's theorem that

$$\theta(n) = 2^n + n - 1.$$

The only minimum 1-coverings are $\alpha\beta$ and $\beta\alpha$. The only minimum 2-covering which starts with $\alpha\beta$ is $\alpha\beta\beta\alpha\alpha$. If we wish a minimum 3-covering which starts with this 5-block we may continue with β or α, but we must then continue with $\alpha\beta$ or $\beta\alpha\beta$ respectively, thereby obtaining the blocks

$$(\alpha\beta\beta\alpha\alpha)\beta\alpha\beta, \qquad (\alpha\beta\beta\alpha\alpha)\alpha\beta\alpha\beta.$$

Further continuation subject to the condition that no 3-block repeats is impossible. Since $\theta(3) = 10$, there exists no transitive ray for which

$$\phi(n) = \theta(n) \qquad (n = 1, 2, 3).$$

It is conceivable that there might exist a transitive ray for which $\phi(n) = \theta(n)$ for all values of n exceeding some fixed integer m. That this is impossible is shown by the following previously unpublished theorem of R. Oldenburger.

THEOREM 11.1. *If the generating symbols reduce to two free symbols* α *and* β *there exists no transitive ray whose ergodic function* $\phi(n)$ *equals the covering index* $\theta(n)$ *for two successive values of* $n > 1$.

The proof as given by Oldenburger is essentially as follows.

We assume the theorem false. There then exists a minimum n-covering $H(n)$, $n > 2$, which starts with a minimum $(n-1)$-covering $H(n-1)$. No n-block repeats in $H(n)$, nor $(n-1)$-block in $H(n-1)$. We shall arrive at a contradiction.

Since α and β can be interchanged throughout it is no essential restriction to assume that in $H(n-1)$ the $(n-1)$-block β^{n-1} occurs before α^{n-1}. We shall then show that $H(n-1)$ has the form

$$H(n-1) = [\beta^{n-1}\alpha \cdots \beta\alpha^{n-1}], \tag{11.1}$$

and that $H(n)$ has the form

$$H(n) = H(n-1)[\alpha\beta \cdots \alpha\beta^{n}]. \tag{11.2}$$

The block of $H(n)$ which remains when $H(n-1)$ is removed will be denoted by J. To establish the correctness of the representations (11.1) and (11.2) one reasons as follows.

The block α^n appears in $H(n)$. It cannot appear wholly in $H(n-1)$, for this would imply repetition of α^{n-1} in $H(n-1)$, and thus $\beta\alpha^n$ appears in $H(n)$. Since β^{n-1} precedes α^{n-1} in $H(n-1)$, $\beta\alpha^{n-1}$ appears in $H(n-1)$. Since $\beta\alpha^{n-1}$ appears just once in $H(n)$, it must appear as the initial n-block of $\beta\alpha^n$, and since this latter block is not wholly in $H(n-1)$, $H(n-1)$ ends with $\beta\alpha^{n-1}$ and J begins with α. This α is of course followed by β.

The block β^n is in $H(n)$ and not wholly in $H(n-1)$. Since J begins with α, β^n must appear wholly in J, and hence $\alpha\beta^{n-1}$ and $\alpha\beta^n$ appear in J. It follows that β^{n-1}, which appears in $H(n-1)$, must appear at the left of $H(n-1)$; otherwise it would be preceded by α, and $\alpha\beta^{n-1}$ would appear twice in $H(n)$. Hence $H(n-1)$ begins with $\beta^{n-1}\alpha$. If J does not end with β^n, $\beta^n\alpha$ appears in J, and $\beta^{n-1}\alpha$ appears both in J and $H(n-1)$. Hence J ends with $\alpha\beta^n$ and the representations (11.1) and (11.2) are correct.

We now make use of the assumption that $n > 2$. Then $n-2 > 0$ and $\alpha\beta^{n-2}$ occurs in $H(n-1)$. This block cannot be followed by β without repetition of β^{n-1} in $H(n-1)$. Hence $\alpha\beta^{n-2}\alpha$ appears in $H(n-1)$. But $\beta^{n-2}\alpha$ already appears at the beginning of $H(n-1)$. Hence $H(n)$ cannot start with a minimum $(n-1)$-covering, and the proof of the theorem is complete.

We turn to the case where the only condition on the generating symbols is the restriction of inverses. This is the simplest case of strict dynamical significance. It arises in the case of surfaces of negative curvature with boundaries. Cf. Morse [1]. In this case one has a set of generating symbols of the form

$$a_1, \cdots, a_p;\ a_1^{-1}, \cdots, a_p^{-1} \qquad (p > 1), \tag{11.3}$$

with the restriction that successive symbols shall not be inverses. We shall prove the following theorem by suitably modifying Martin's proof.

THEOREM 11. 2. *For generating symbols of the form* (11. 3) *subject to no condition beyond the restriction of inverses there exists a minimum r-covering in which no r-block repeats.*

The following rule for the formation of a minimum r-covering $E(r)$ was first given by Martin for the case of free generating symbols $c_1, \cdots, c_n$. Start with c_1^{r-1} and continue successively adjoining the generating symbol of highest index consistent with the condition that no r-block be repeated. This will lead to a minimum r-covering in the case of free symbols c_i. For the case of inverses we adjoin the condition that successive symbols shall not be inverses. With this condition added the process will not always lead to a minimum r-covering. We are able however to prove the following lemma.

LEMMA 11. 1. *A necessary and sufficient condition that the application of the Martin rule, subject to the restriction of inverses, shall lead to a minimum r-covering is that c_1 and c_2 shall not be inverses in the set $c_1, \cdots, c_n$ of generating symbols.*

Let M_r be the block of maximum length obtained by applying the Martin rule subject to the restriction of inverses. If A is any admissible $(r-1)$-block, the r-blocks

$$Ac_n, \cdots, Ac_1 \tag{11.4}$$

appear in M_r, if at all, in the order written. One of the blocks (11. 4) is inadmissible since the last symbol in A cannot be followed by its inverse. If Ac_1 is in M_r, all other admissible blocks in (11. 4) are in M_r. If Ac_1 is not admissible, Ac_2 is admissible and is in M_r only if all of its predecessors in (11. 4) are in M_r. We set $E = c_1^{r-1}$. We continue with a proof of (a) and (b).

(a). *The block M_r ends with E.* If M_r ended with an $(r-1)$-block $A \neq E$, M_r could not begin with A, since M_r begins with E. A would then appear in M_r at most $n-1$ times; otherwise some one of the $n-1$ admissible blocks c_iA would appear twice. Following A's last appearance we can continue M_r with c_2 (or c_1 if Ac_2 is inadmissible), contrary to the hypothesis that M_r ends with A. Hence M_r ends with E.

(b). *The block M_r ends with Ec_1.* Suppose M_r ends with an r-block $B \neq Ec_1$. Set $B = bE$ in accordance with (a). We are supposing that $b \neq c_1$. The block E appears n times in M_r; otherwise M_r could be continued. Hence Ec_1 appears in M_r but not at the end. But Ec_1 cannot be continued in accordance with the Martin rule. From this contradiction we infer the truth of (b).

We shall now make use of the assumption that c_1 and c_2 are not inverses and show that M_r contains all r-blocks.

We first observe that each admissible block Ec_j appears in M_r since M_r ends with Ec_1. Hence M_r contains n copies of E. It follows that M_r contains each admissible block c_jE. Let $B = b_1 \cdots b_r$ be an arbitrary admissible r-block. We shall show that M_r contains B.

Suppose that M_r does not contain B. Set $b_2 \cdots b_r = D$. Then $D \neq E$ by virtue of (b). Hence D appears at most $n-2$ times in M_r; otherwise each admissible block c_jD would appear in M_r, including B. Hence Dc_1 does not appear in M_r (or Dc_2, if Dc_1 is inadmissible). We now apply the same reasoning to Dc_1 (or Dc_2) that we have applied to B and infer that $b_3 \cdots b_r c_1 c_1$ (or $b_3 \cdots b_r c_2 c_1$) does not appear in M_r. Continuing we arrive at the conclusion that $c_1{}^r$ does not appear in M_r, contrary to (b).

The condition of the lemma is accordingly sufficient.

To prove the condition necessary we suppose that $c_1 = c_2{}^{-1}$. If M_r were an r-covering $c_2{}^r$ would appear in M_r, but not in the last position, by virtue of (a). But $c_2{}^r$ cannot be continued without violating the Martin rule, or else the condition on inverses. Hence the condition is necessary, and the proof of the lemma is complete.

Theorem 11. 2 is an immediate consequence of the lemma.

If there are n generating symbols subject merely to the restriction of inverses, there are $n(n-1)^{r-1}$ different admissible r-blocks so that the r-th covering index $\theta(n)$ has the value

$$\theta(r) = n(n-1)^{r-1} + r - 1.$$

With this understood we shall prove the following theorem.

THEOREM 11. 3. *When there are n generating symbols subject merely to* *the restriction of inverses, each ergodic function* $\psi(r)$ *is at least* $\theta(r)$ *and there exists an ergodic function* $\phi(r)$ *such that*

$$\phi(r) < \left(\frac{n-1}{n-2}\right) n(n-1)^{r-1} < \frac{n-1}{n-2}\,\theta(r). \tag{11.5}$$

Let H_r be a minimum r-covering and let H^*_r be the block obtained from H_r by omitting the first $r-1$ cymbols c_1. The ray

$$R = H_1 H^*_2 H^*_3 \cdots$$

is admissible and transitive. We note in particular that H^*_r is preceded by c_1^{r-1} in R. For this ray $\phi(r)$ is at most the length of the block $H_1 H^*_2 \cdots H^*_r$, that is

$$\phi(r) \leqq \sum_{m=1}^{r} n(n-1)^{m-1}.$$

Hence

$$\phi(r) \leqq n\left[\frac{1-(n-1)^r}{1-(n-1)}\right] = \left(\frac{n-1}{n-2}\right) n \left[(n-1)^{r-1} - \frac{1}{n-1}\right],$$

and (11.5) follows at once.

We conclude this section by proving the following theorem.

THEOREM 11.4. *If* $N(r)$ *is the number of admissible* r*-blocks under our general conditions of admissibility, there exists an ergodic function* $\phi(r)$ *such that*

$$\phi(r) < (r+1)N(r), \tag{11.6}$$

while each ergodic function is at least $N(r) + r - 1$.

We begin by proving the following.

(a). *There exists an* r*-covering of length less than* $(r+1)N(r)$.

Statement (a) is true if $r = 1$. For if $a_1, \cdots, a_n$ is the set of generating symbols, there exists a 1-covering of the form $a_1 e_1 a_2 e_2 \cdots e_{n-1} a_n$ provided the symbols e_i are suitably chosen. This 1-covering has the length $2n - 1 < 2N(1)$. We assume then that A_r is an r-covering of length less than $(r+1)N(r)$. We shall show how to form an $(r+1)$-covering of length less than $(r+2)N(r+1)$.

There are $N(r)$ distinct r-blocks in A_r and hence at least $N(r) - 1$ distinct $(r+1)$-blocks. There remain at most

$$\mu = N(r+1) - N(r) + 1$$

$(r+1)$-blocks, say $B_1, \cdots, B_\mu$ not in A. The block

$$A_{r+1} = A_r e_1 B_1 e_2 B_2 e_3 \cdots e_\mu B_\mu,$$

in which the symbols e_i are chosen so as to make A_{r+1} admissible, is an $(r+1)$-covering. Let s_i be the length of A_i. We see that

$$s_{r+1} = s_r + (r+2)\mu.$$

By virtue of the hypothesis that $s_r < (r+1)N(r)$ we infer that

$$\begin{aligned} (11.7) \qquad s_{r+1} &< (r+1)N(r) + (r+2)[N(r+1) - N(r) + 1] \\ &= (r+2)N(r+1) - N(r) + r + 2. \end{aligned}$$

But for $r > 1$,

$$N(r) \geqq 2^r \geqq r+2,$$

so that (11.7) yields the relation

$$s_{r+1} < (r+2)N(r+1),$$

and the proof of (a) is complete.

In proving (a) we have defined r-coverings A_r, $r = 1, 2, \cdots$, such that A_r is an initial block of A_{r+1}. A ray whose first s_r symbols are those of A_r for each r will satisfy (11.6). The final statement of the theorem is immediate.

12. Category and sets of transitive rays. In this section we shall determine the category of the set of transitive rays or of sets of relatively transitive rays. More particularly, we shall determine the category of various subsets of transitive rays defined by limitations on the ergodic functions. In this way we obtain an idea of the distribution of transitive rays. A similar use can be made of measure theory (cf. § 13), but the development of this phase of the theory is deferred to a later paper.

In the category theory, the space used will be the space S of admissible rays with the Baire metric. Cf. § 4. Recall that the space S is compact, perfect, and totally disconnected. Let H be a perfect subset of S. A set A is of the *first category* relative to H if it is the sum of an at most enumerable set of nowhere dense (relative to H) subsets of H. These subsets will be called the components of A. The components of A may vary from the extreme case where they are all null to the case where they are perfect. The complement $H - A$ of a set A of the first category is termed a *residual set* of the second category. Such a set is everywhere dense in H and has the power of the continuum. Cf. Hobson [1], p. 132.

Let R be a given ray. The rays obtained by deleting a finite initial block of R will be said to be *based* on R, those obtained by the inverse process of adding a finite block to R will be termed *extensions* of R. The rays which are limit rays in the space S of the set of rays based on R will be called *limit* rays of R. A ray R which is a limit ray of itself will be said to be *transitive relative to its limit rays.* It is thereby necessary and sufficient that R be non-periodic and that every block of R appear infinitely many times in R. The limit rays of a relatively transitive ray R form a closed set H. This set is identical with its derived set since each ray based on R is a limit ray of the set. The set H is accordingly perfect. Concerning the set H we have the following theorem.

THEOREM 12. 1. *In the set H of limit rays of a relatively transitive ray R, the rays which are transitive relative to H form a residual set in H.*

Let $A_1, A_2, \cdots$ be an enumeration of the blocks of R. Let E_n be the subset of rays of H which contain no copy of A_n. The set E_n is closed and contains no rays based on R. It is nowhere dense in H since the rays based on R are everywhere dense in H. Finally the set S of rays which are transitive relative to H is of the form

$$S = H - \Sigma E_n,$$

and the theorem follows from the definition of a residual set in H.

The above sets E_n may all be null, in fact are all null if R is a subray of a non-periodic recurrent trajectory. In the case where R is transitive, that is transitive relative to the set of all admissible rays, the sets E_n are never null as the following lemma shows.

LEMMA 12. 1. *The set K of rays which exclude an admissible r-block B for which $r > 2$ is closed, nowhere dense, and of the power of the continuum.*

It is clear that K is closed. That K is nowhere dense in S follows from the existence of a transitive ray R. For the rays based on R are everywhere dense in S and transitive. Since K is closed, it must then be nowhere dense.

It remains to prove that K is of the power of the continuum. To that end we consider the admissible rays

(12. 1) $\alpha^{r_1}\beta\alpha^{r_2}\beta\alpha^{r_3}\beta \cdots, \qquad r_i = 2 \text{ or } 3,$

formed from the free symbols α and β. This set is in one-to-one correspondence with the dyadic decimals

$$.r_1r_2r_3\cdots, \qquad r_i = 2 \text{ or } 3,$$

and hence is of the power of the continuum.

Similarly, the set

$$\beta^{r_1}\alpha\beta^{r_2}\alpha\beta^{r_3}\cdots, \qquad r_i = 2 \text{ or } 3, \tag{12.2}$$

consists of admissible rays and is of the power of the continuum.

A ray of the form (12.1) has no 3-blocks in common with a ray of the form (12.2). Since the length of B is at least 3, it cannot appear in a ray of (12.1) and a ray of (12.2). Hence at least one of the sets (12.1) and (12.2) is contained in K. Thus K must be of the power of the continuum.

The set of non-transitive rays is of the first category. But much more than this can be said, as the following theorem shows.

THEOREM 12.2. *The set Σ of non-transitive rays is everywhere dense in S and is the sum of countably many nowhere dense, closed sets, each of which is of the power of the continuum.*

That the set Σ is everywhere dense follows from the fact that an arbitrary block B and a sequence A of the symbols α can be combined in the form BeA to form a non-transitive ray. Rays of this form are everywhere dense in S. Let $A_1, A_2, \cdots$ be an enumeration of admissible r-blocks with $r > 2$. The set E_n of rays which exclude A_n is closed, nowhere dense and of the power of the continuum, in accordance with Lemma 12.1. The set Σ is the sum of the sets E_n, and the proof is complete.

THEOREM 12.3. *If $f(n)$ is an arbitrary positive function of the positive integers n, the set F of transitive rays with ergodic function $\phi(n) > f(n)$, for all values of n, has the power of the continuum.*

Let $A_1, A_2, \cdots$ be an enumeration of admissible blocks consistent with their lengths. Let $r_1, r_2, \cdots$ be an arbitrary sequence of positive integers and let $c_1c_2\cdots$ be an arbitrary sequence of the symbols α and β. We introduce the block

$$B_i = \alpha^{r_i}c_i\alpha^{r_i}e_iA_i \qquad (i = 1, 2, \cdots),$$

choosing e_i so that B_i is admissible. We then introduce the ray

$$E = B_1 b_1 B_2 b_2 \cdots,$$

choosing the symbols b_i so that the ray E is admissible. It is clear that E is transitive.

Let $\phi(n)$ be the ergodic function of E. We choose r_1 so that $\phi(n) > f(n)$ for $n = 1, 2, 3$. Let B be the initial $f(4)$-block of E. If r_2 is chosen sufficiently large, blocks $A_1, A_2, \cdots, A_p$ entering into the definition of E will all be 1-blocks. In fact, if r_2 is sufficiently large only A_1 will enter into the definition of B. If A_i is a 1-block, $e_i A_i b_i$ is a 3-block so that B cannot contain β^4. With r_2 so chosen, $\phi(4) > f(4)$.

Proceeding inductively with $m > 2$ we choose r_m so large that in the initial $f(m+2)$-block C of E, at most $(m-1)$-blocks A_i are employed. The corresponding blocks $e_i A_i b_i$ are at most $(m+1)$-blocks so that β^{m+2} does not appear in C. Hence

$$\phi(m+2) > f(m+2) \qquad (m = 3, 4, \cdots). \tag{12.3}$$

Thus $\phi(n) > f(n)$ for all values of n.

The above symbols c_i can be chosen arbitrarily from the set α, β the r_i's remaining fixed. For fixed r_i's there are then as many different rays E as there are different sequences $c_1 c_2 \cdots$, that is a number equal to the power of the continuum.

The proof of the theorem is complete.

The set F of the theorem is not closed. The set of transitive rays not in F is open in the set of all transitive rays so that F is everywhere dense only if it includes all transitive rays.

THEOREM 12.4. *Given an arbitrary function $f(n)$, the set ω of transitive rays R with ergodic functions $\phi(n) \leqq f(n)$ for n exceeding some integer dependent on R is a set of the first category.*

Let D_m be the set of transitive rays for which $\phi(n) \leqq f(n)$ for n exceeding m. The set D_m is closed. But we have seen in Theorem 12.2 that the set of intransitive rays is everywhere dense. Hence D_m is nowhere dense. But $\omega = \Sigma D_m$, and the proof is complete.

Let $H(n)$ be a block of minimum length $\theta(n)$ covering all admissible n-blocks. We have termed $\theta(n)$ the n-th covering index. As we have seen in § 11, there is in general no transitive ray with ergodic function $\phi(n) = \theta(n)$. We have however the following theorem.

THEOREM 12.5. *The set G of transitive rays with ergodic function $\phi(n)$ such that*

$$\lim_{n\to\infty}\inf.\frac{\phi(n)}{\theta(n)}=1, \tag{12.4}$$

where $\theta(n)$ is the n-th covering index, is a residual set of the second category.

We begin with a proof of the following.

(a). *The set G is not empty.*

To prove (a) let δ_n be a sequence of positive constants converging to zero. We shall consider a ray of the form

$$R=H(r_0)e_1H(r_1)e_2H(r_2)\cdot\cdot\cdot,$$

where $H(n)$ is defined as above and the symbols e_i are chosen so that R is admissible. The integers r_i are chosen as follows. Let $r_0=1$. Proceeding inductively we suppose that r_{n-1} is defined and that the block

$$H(r_0)e_1H(r_1)e_2\cdot\cdot\cdot e_{n-1}H(r_{n-1})e_n$$

has the length m_n. We then choose r_n so large that

$$\frac{\theta(r_n)+m_n}{\theta(r_n)}<1+\delta_n. \tag{12.5}$$

If $\phi(n)$ is the ergodic function of the resulting ray R, it is clear that

$$\phi(r_n)\leqq\theta(r_n)+m_n,$$

since the initial block of length $\theta(r_n)+m_n$ in R contains $H(r_n)$ and hence all r_n-blocks. The function $\phi(n)$ will satisfy (12.4) by virtue of (12.5), and the proof of (a) is complete.

If A is an arbitrary admissible r-block, the ray $X=AeR$ is admissible if e is suitably chosen. Moreover the ray R and the ray X have ergodic functions ϕ and ψ respectively such that

$$\theta(n)\leqq\psi(n)\leqq\phi(n)+r+1.$$

Hence X satisfies the condition (12.4) and is in G. We conclude that the rays of G are everywhere dense.

To come to a proof of the theorem let m be a positive integer and p a rational number >1. The rays for which the transitivity index $\psi(m)$ satisfies the condition

$$\psi(m) > \theta(m)p$$

form a closed set $D_m{}^p$. The intersection

$$E_q{}^p = D_q{}^p D^p{}_{q+1} D^p{}_{q+2} \cdots$$

is closed. The complement of G is the sum of the sets $E_q{}^p$ as q ranges over all positive integers and p ranges over all rational numbers > 1. But G is everywhere dense so that $E_q{}^p$ is nowhere dense, and the proof of the theorem is complete.

THEOREM 12.6. *The set Z of transitive rays with ergodic functions $\phi(n)$ such that*

$$\lim_{n\to\infty}\text{inf.}\frac{\phi(n)}{\theta(n)} > 1$$

is everywhere dense, has the power of the continuum, and is of the first category.

The set Z includes the set of transitive rays for which $\phi(n) \geqq 2\theta(n)$, and so has the power of the continuum in accordance with Theorem 12.3. The set Z is a subset of the complement of the residual set G of Theorem 12.5 and so is of the first category. It remains to show that Z is everywhere dense.

If R is a ray of Z, any ray R' based on R is again transitive. If the initial symbol of R' is the $(r+1)$-st symbol of R, the ergodic functions $\psi(n)$ and $\phi(n)$ of R' and R, respectively, satisfy the conditions

$$\psi(n) \geqq \phi(n) - r, \qquad \lim_{n\to\infty}\text{inf.}\frac{\psi(n)}{\theta(n)} > 1.$$

Hence R' belongs to Z. But the rays R' based on the transitive ray R are everywhere dense so that Z is everywhere dense as stated.

The proof of the theorem is complete.

13. Trajectories with the separation property.

The results of this section will form the basis of a more extensive treatment to be presented later. No proofs will be given in this place.

Let t be a trajectory involving two generating symbols α and β. By the α-length (β-length) of a block B of t we shall mean the number of α's (β's) in B. The trajectory t will be said to have the *separation property* and be of *type S* if the following holds.

The α-lengths of any two n-blocks of t differ by at most 1.

The preceding condition implies the corresponding condition for β-lengths.

Trajectories of type S arise in many ways. In particular let

$$\frac{d^2y}{dx^2} + f(x)y = 0 \tag{13.1}$$

be a differential equation in which $f(x)$ is continuous and has the period 1. Let $u(x)$ be a solution of (13.1) which vanishes at least once, but not identically. We are concerned with the distribution of the zeros of $u(x)$. Without loss of generality we can suppose that none of these zeros occur at the points $x = n$, where n is an integer, positive, negative or zero. For the zeros of $u(x)$ are countable, and for our purposes any transformation of the form $\bar{x} = x + c$ is admissible. With this understood let a trajectory t be defined as follows. Let each point on the x-axis at which $u(x) = 0$ be replaced by the symbol α, and each point $x = n$ be replaced by the symbol β. The resulting trajectory t will be of type S. With trivial exceptions every trajectory of type S can be obtained in this way. The study of trajectories of type S is thus a study of the zeros of solutions of equations of the form (13.1).

Let $\alpha_n(\beta_n)$ denote the α-length (β-length) of some n-block of a trajectory t of type S. It can be shown that there exists a number θ, $0 \leqq \theta \leqq 1$, such that

$$\lim_{n\to\infty} \frac{\alpha_n}{n} = \theta, \qquad \lim_{n\to\infty} \frac{\beta_n}{n} = 1 - \theta.$$

The number θ will be termed the *rotation number* of t. Cf. Poincaré [1].

The trajectory is not uniquely determined by the rotation number. If θ is irrational, there are infinitely many trajectories of type S having θ as rotation number and these trajectories are all recurrent. Thus trajectories of type S are in general recurrent. A simple exceptional example is the trajectory in which all the symbols are α except for one β, in which case the rotation number is 1.

Given θ, $0 \leqq \theta \leqq 1$, the recurrent trajectories of type S which have θ as rotation number form a minimal set which consists of a single periodic trajectory if and only if θ is rational. Since all of the trajectories of one of these minimal sets have the same recurrency function, it can be denoted by $R_\theta(n)$. The behavior of $R_\theta(n)$ varies widely with the choice of θ, but in all cases

$$\liminf_{n\to\infty} \frac{R_\theta(n)}{n} \leqq 4.$$

There exist irrational values of θ, and thus non-periodic recurrent trajectories such that

$$\lim_{n\to\infty}\text{sup.}\ \frac{R_\theta(n)}{n} \leqq 6.5.$$

On the other hand, given an arbitrary function $f(n)$, there exist irrational values of θ such that $R_\theta(n) > f(n)$ for infinitely many values of n.

If certain exceptional θ-sets of (Lebesgue) measure zero are excluded, $R_\theta(n)$ displays great regularity in its asymptotic behavior. We can state that, except for a θ-set of measure zero,

$$\lim_{n\to\infty}\text{sup.}\ \frac{R_\theta(n)}{n} = +\infty,$$

while if $c > 1$, except for a θ-set of measure zero,

$$\lim_{n\to\infty}\text{sup.}\ \frac{R_\theta(n)}{n^c} \leqq 1.$$

14. Typical open questions. Let R be a ray of the form $c_0c_1c_2\cdots$ and let t be a trajectory containing R as a subray. Among the positive limit trajectories of t are found one or more distinct minimal sets of trajectories. Cf. Theorem 6.1. Suppose v_1 of these sets are periodic trajectories, and v_2 of these sets general minimal sets. The number v_1 is at most aleph null, and v_2 is at most aleph. That v_2 may equal aleph will be shown by an example in a later paper. The numbers v_1 and v_2 are uniquely determined by R and will be called the *limit indices* of R of the first and second types respectively. Our first question is as follows.

(1). What are the limit indices of algebraic numbers, regarding decimal representations as rays? Are these rays ever transitive or recurrent? What are the answers to similar questions concerning π and e?

(2). Let t be a non-periodic recurrent trajectory. In all known examples $R(n)/n$ fails to converge. Is this true in general?

(3). If t is a non-periodic recurrent trajectory, $R(n) \geqq kn$ where $k \geqq 2$. Cf. Theorem 7.5. What is the least value of k such that $R(n) \geqq kn$ for each non-periodic recurrent trajectory?

(4). In the case of general admissibility certain blocks and finite sequences of blocks are prohibited. Knowing these prohibitions, is it possible to give a constructive rule for the formation of a minimum block covering all admissible r-blocks, and what is the most general such r-covering?

(5). Does there exist an ergodic function $\phi(n)$ such that the ratio of $\phi(n)$ to $\theta(n)$ tends to the limit 1? Cf. Theorem 12.5.

(6). Is a theorem similar to that of Oldenburger true in case the generating symbols suffer the restriction of inverses, or in the general case?

(7). What is the most general minimal set with a given recurrency function $R(n)$? With permutation index $P(n)$? With $P(n)$ and $R(n)$ both given? These questions should be answered first when there are just two generating symbols.

These questions seem to point to an algebraic analysis. Questions more intimately connected with measure theory, of the analysis of space forms will be presented in a later paper.

THE INSTITUTE FOR ADVANCED STUDY,
AND
BRYN MAWR COLLEGE.

BIBLIOGRAPHY.

ARTIN, E.

1. "Ein mechanisches System mit quasiergodischen Bahnen," *Abhandlungen aus dem Mathematischen Seminar der Hamburgischen Universität*, Bd. 3 (1924), pp. 170-175.

BESICOVITCH, A. S.

1. *Almost periodic functions*, Cambridge, University Press (1932).

BIRKHOFF, G. D.

1. "Quelques théorèmes sur le mouvement des systèmes dynamiques," *Bulletin de la Société Mathématique de France*, t. 40 (1912), pp. 303-323.
2. "Dynamical systems," *American Mathematical Society Colloquium Publications*, vol. 9 (New York, 1927).
3. "Nouvelles recherches sur les systèmes dynamiques," *Memoriae Pont. Acad. Scient. Novi Lyncaei* (Série 3), t. 1, pp. 85-216.
4. "Sur le problème restreint des trois corps" (Second Mémoire), *Annali della R. Scuola Normale Superiore di Pisa* (Serie II), t. 5 (1936), pp. 1-42.

HADAMARD, J.

1. "Les surfaces à courbure opposées et leur lignes géodésiques," *Journal de Mathématiques pures et appliquées* (5th series), t. 4 (1898), pp. 27-73.

HAUSDORFF, F.

1. *Mengenlehre,* Leipzig, Walter de Gruyter (1927).

HEDLUND, G. A.

1. "On the metrical transitivity of the geodesics on closed surfaces of constant negative curvature," *Annals of Mathematics,* vol. 35 (1934), pp. 787-808.
2. "A metrically transitive group defined by the modular group," *American Journal of Mathematics,* vol. 57 (1935), pp. 668-678.
3. "Two-dimensional manifolds and transitivity," *Annals of Mathematics,* vol. 37 (1936), pp. 534-542.

HILBERT, D.

1. "Über Flächen von konstanter Gaussscher Krümmung," *Transactions of the American Mathematical Society,* vol. 2 (1901), pp. 87-99.

HOBSON, E. W.

1. *The Theory of Functions of a Real Variable,* Cambridge, University Press (1921).

KOEBE, P.

1. "Riemannsche Mannigfaltigkeiten und nicht euklidische Raumformen," *Sitzungsberichte der Preussischen Akademie der Wissenschaften* (1927), pp. 164-196; (1928), pp. 345-442; (1929), pp. 414-457; (1930), pp. 304-364, 505-541; (1931), pp. 506-534.

MARTIN, M. H.

1. "A problem in arrangements," *Bulletin of the American Mathematical Society,* vol. 40 (1934), pp. 859-864.
2. "The ergodic function of Birkhoff," *Duke Mathematical Journal,* vol. 3 (1937), pp. 248-278.

MORSE, M.

1. "A one-to-one representation of geodesics on a surface of negative curvature," *American Journal of Mathematics,* vol. 43 (1921), pp. 33-51.
2. "Recurrent geodesics on a surface of negative curvature," *Transactions of the American Mathematical Society,* vol. 22 (1921) pp. 84-100.
3. "A fundamental class of geodesics on any closed surface of genus greater than one," *Transactions of the American Mathematical Society,* vol. 26 (1924), pp. 25-61.
4. "Instability and transitivity," *Journal de Mathématiques pures et appliquées* (9th series), t. 14 (1935), pp. 49-71.

MYRBERG, P. J.

1. "Einige Anwendungen der Kettenbrüche in der Theorie der binären quadratischen Formen und der elliptischen Modulfunktionen," *Annales Acad. Sc. Fennicae,* t. 23 (1924).
2. "Ein Approximationssatz für die Fuchsschen Gruppen," *Acta Mathematica,* t. 57 (1931), pp. 389-409.

NIELSEN, J.

1. "Om geodaetiske Linier i lukkede Mangfoldigheder med konstant negativ Krumning," *Matematisk Tidskrift* B, (1925), pp. 37-44.
2. "Untersuchungen zur Topologie der geschlossenen zweiseitigen Flächen," *Acta Mathematica*, t. 50 (1927), pp. 189-358.

POINCARÉ, H.

1. "Sur les courbes définies par les équations différentielles," *Journal de Mathématiques pures et appliquées* (Série 4), t. 1 (1885), pp. 167-244.

ROBBINS, H. E.

1. "On a class of recurrent sequences," *Bulletin of the American Mathematical Society*, vol. 43 (1937), pp. 413-417.

PERIODIC ORBITS IN THE RESTRICTED PROBLEM OF THREE BODIES IN THEIR RELATION TO HILL'S WORK ON THE MOTION OF THE MOON.

By Elis Strömgren.

The first volume of the *American Journal of Mathematics* (1878) contains two papers by G. W. Hill that have become of fundamental importance in the theory of the motion of the moon, and which, besides, contain a few remarks that may well be looked upon as the starting point of important chapters of that special case of the problem of three bodies known as "The restricted problem of three bodies," or "Problème restreint."

The Problème restreint was first defined by Jacobi: Three bodies are given, two of which have finite masses and are moving around each other in circular orbits according to the laws of the problem of two bodies. The problem is to study the motion of the third body, assumed to have an infinitely small mass, when moving in the plane of the two finite masses under the influence of the attraction from these two masses. With regard to the mass ratio, it should be noticed that Poincaré in his investigations of the restricted problem always assumes that one of the two finite masses is very small relative to the other, thus reducing the problem to a perturbational problem, which can be solved with the aid of series expressions in powers of a small quantity, but which on the other hand enables us to study the nature of the motion in the neighbourhood of known exact solutions only.

In the work carried out at the Copenhagen Observatory on the other hand, the two finite masses have been assumed to be equal, and the scope of the problem is essentially enlarged, it being desired to follow all investigated classes of orbits throughout the whole of their evolution. This problem, however, not being a simple perturbational problem, could only be solved with the aid of *numerical* methods of integration.

The cusps in Hill's variational curves. Plate III in vol. 1 of the *American Journal of Mathematics* gives (cf. Fig. 1 in this paper) a series of lunar orbits computed by Hill in the idealized variational problem, in which the parallactic factor is put equal to zero. In the diagram we see four variational curves from 1 to 4, of which the last has two symmetrical cusps, one above, and one below.

With regard to the orbits in a continued evolution of these variational curves Hill says as follows on p. 259 (*Collected Works*, I, 334): "It is certain that the orbits, if they do exist, do not intersect the line of quadratures, and that the moons describing them would make oscillations to and fro, never departing as much as 90° from the point of conjunction or of opposition." This was only a conjecture, and it was shown to be false by Poincaré in *Méthodes Nouvelles*, I, 106 ff. (1892). Here Poincaré shows by a simple mathematical procedure that the cusps in Hill's variational curves in the further evolution simply become *loops*. Later (1907) in the *Astronomische Nachrichten* No. 4155, the author of this article drew attention to the fact that, in the restricted problem, loops and cusps can exist at any point in the plane, with the exception of certain special points. With regard to these loops and cusps, the author found a general theorem, to which we shall return.

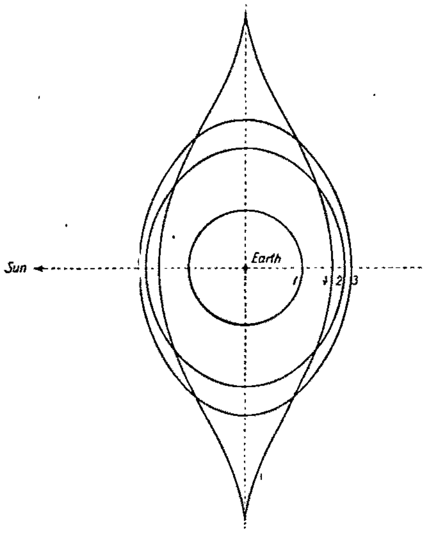


Fig. 1.
Hill's variational curves.

Hill's division of the plane of motion into permitted and forbidden regions. In the first of the two papers in the first volume of the *American Journal of Mathematics* quoted above, Hill, with the aid of the so-called Jacobian integral, makes his famous investigation of the permitted and the forbidden regions in the plane of motion. The division is made with the aid of the so-called curves of zero velocity. These curves are shown in Fig. 2 for the mass ratio of the two finite masses $m_1 = 10m_2$ used by G. H. Darwin in his "Periodic orbits." Plate I shows the curves for the mass ratio $m_1 = m_2$

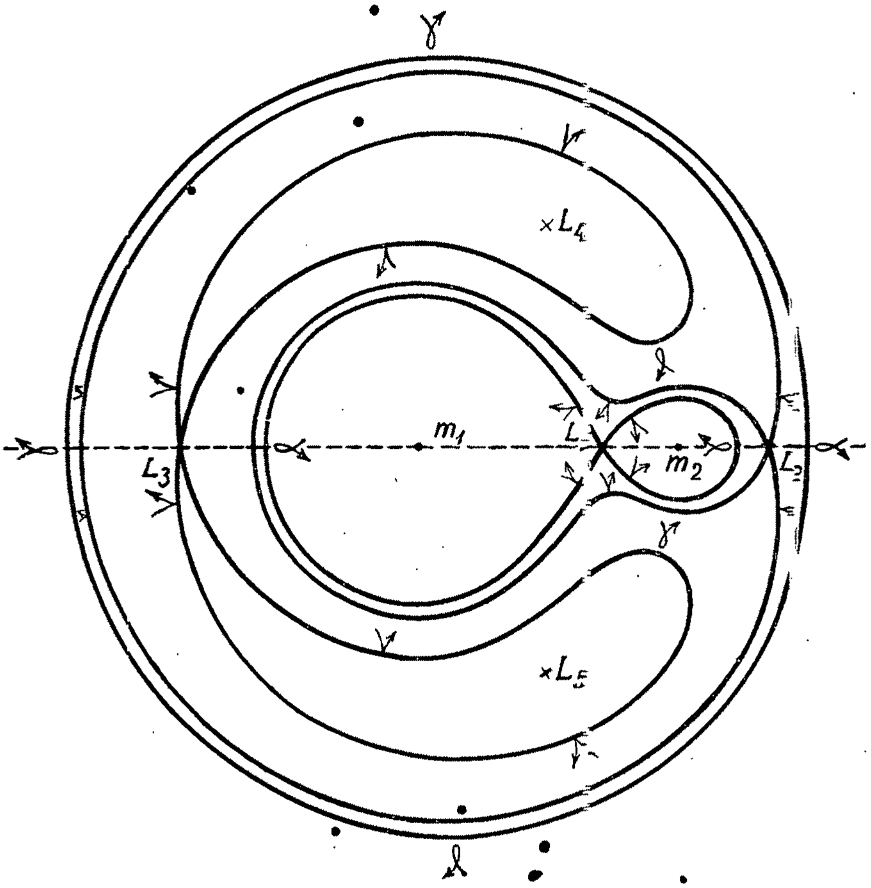


Fig. 2.
Zero velocity-curves in the restricted problem of three bodies, with $m_1 = 10m_2$.

suggested by Thiele and adopted in the work at the Copenhagen Observatory (Plate I is based upon a drawing made by the Moscow astronomer N. Moisseiev. The scale and the values of the so-called Jacobian constant are the same as in the work at the Copenhagen Observatory, cf. for instance Publ. No. 100 of the Copenhagen Observatory, or *Bulletin Astronomique*, Deuxième Série, Tome IX, Fasc. 2, 1935). In both of these drawings we have had cusps and loops drawn especially, to illustrate the general theorem referred to above (*Astronomische Nachrichten* No. 4155): *When, located on that part of the normal of a zero velocity curve which belongs to the permitted region, we look in the direction toward the curve, then the motion of the infinitely small mass in a cusp is always from right to left.* As is seen from Fig. 2 and Plate I the direction of motion in a loop also follows immediately from this theorem. It should be noticed, that in these diagrams, as in Fig. 1 and all the following diagrams, the X-axis is the line from m_1 to m_2, which, in a fixed coördinate system is rotating with uniform velocity. The 5 points of libration are denoted by L_1, L_2, L_3, L_4, L_5 in Fig. 2.

The evolution of a cusp during the evolution of a class of orbits. The programme of the work at the Copenhagen Observatory originally was this: to find all classes of simply-periodic orbits around the finite masses and around the points of libration, and to follow these classes through the whole of their evolution. This programme has at the present time, with one single exception (cf. Publ. 100, pp. 26-27), been completely carried out. As the work progressed, however, this programme was automatically considerably extended, as we were, during the investigations, led to study quite different possibilities of finding periodic solutions in the problem.

Among the classes of orbits on our original programme was, amongst others, the following: It was clear a priori that at great distances from the origin orbits with slow, almost circular motion would exist, which in a fixed coördinate system were *direct* (motion in the same direction as the two finite masses), and—on account of the rotation of the m_1m_2-axis—rapid *retrograde* in the rotating coördinate system. Fig. 3 shows a number of orbits in this class. Far out we have an almost circular orbit ($K = 14.4361$). Then we see an orbit ($K = 12.60816$) of markedly oval shape, and still further inwards we encounter an orbit ($K = 11.76718$) with a marked indentation. This indentation develops in the evolution that follows into a cusp, and then into a loop.

Fig. 4-7 show how the evolution is continued. The evolution ends in the point of libration L_4 *with asymptotic motion into as well as out from this point of libration, on the ingoing branch with an infinite number of steadily decreasing windings, with the velocity converging toward zero,* on the outgoing branch the reverse. Fig. 8 shows the ingoing branch only.

This was an example of the asymptotic motions predicted by Poincaré in *Méthodes Nouvelles* on a theoretical basis, the asymptotic motion being here periodic also.

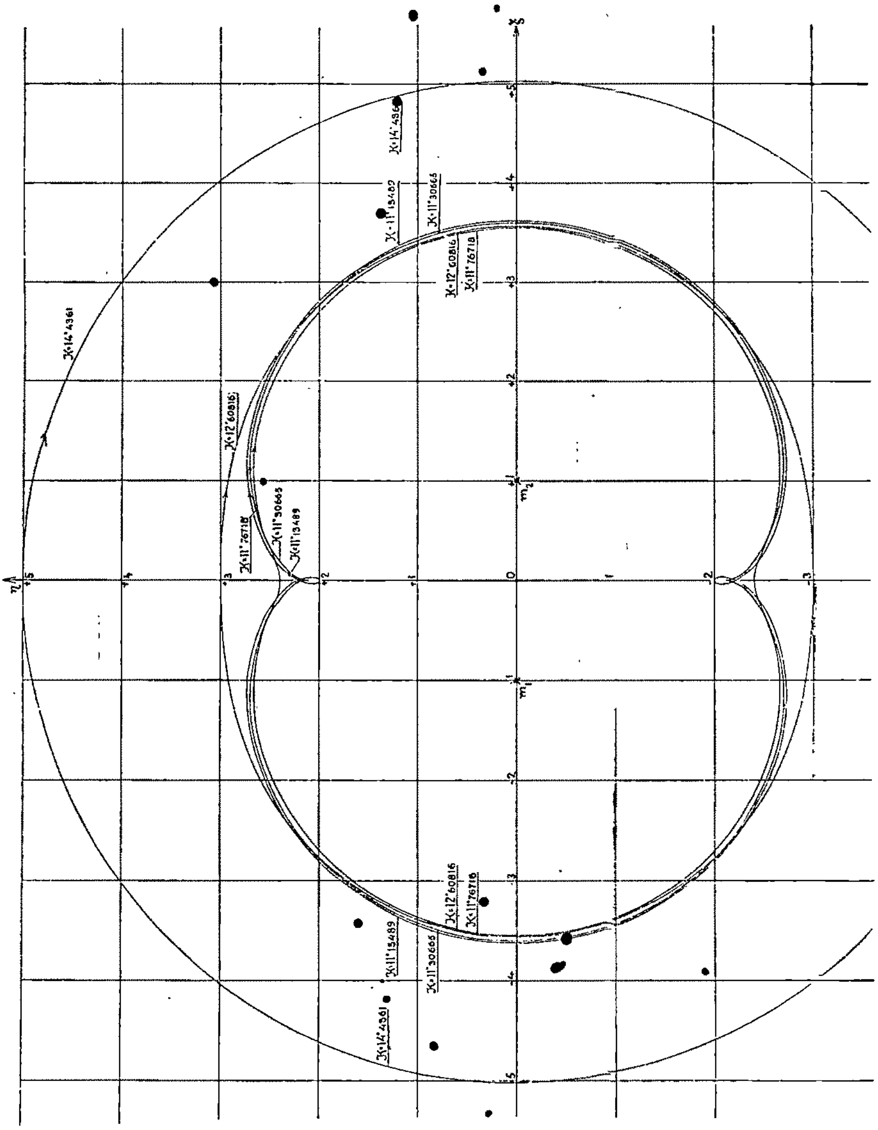


Fig. 3.

A class of periodic orbits in the restricted problem of three bodies.

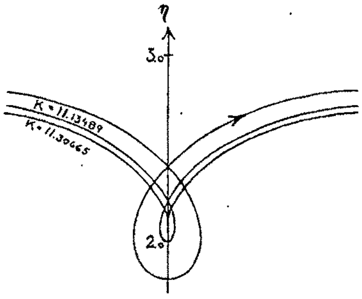


Fig. 4.

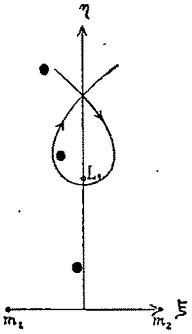


Fig. 5.

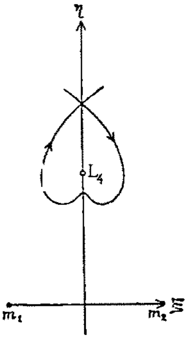


Fig. 6.

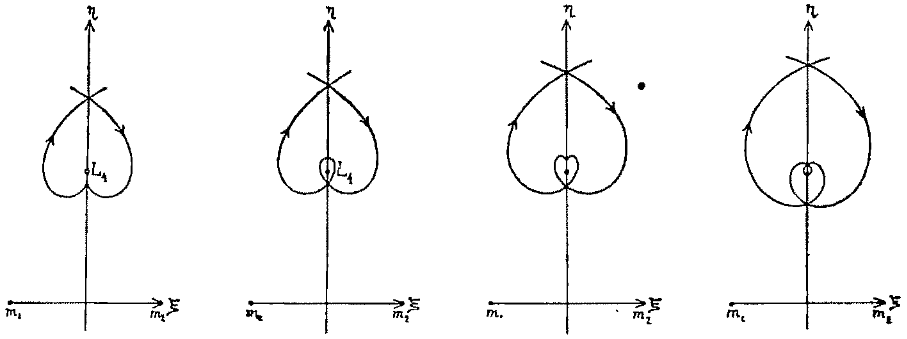


Fig. 7.

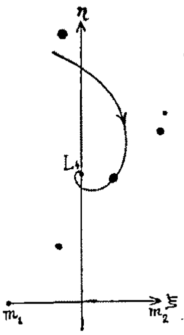


Fig. 8.

The evolution of cusps and loops up to asymptotic motion.

Asymptotic orbits, that are also periodic, with an infinitely long period. We have seen how the class of orbits shown in Fig. 3, which begins at infinity, ends with a periodic orbit of infinite period, asymptotic to the two points of libration L_4 and L_5. This is a special case of a general theorem, that was deduced in an empirical way as a result of our investigations: It has been found that, as far as our experience goes, a class of periodic orbits always, either

I) goes back into itself, so as to form a class closed in itself,
or II) has a natural beginning and a natural ending. Such a natural beginning or natural ending always takes place:
either a) in one of the two finite masses, or in both,
or b) in one of the 5 libration points, or in two of them,
or c) at infinity.

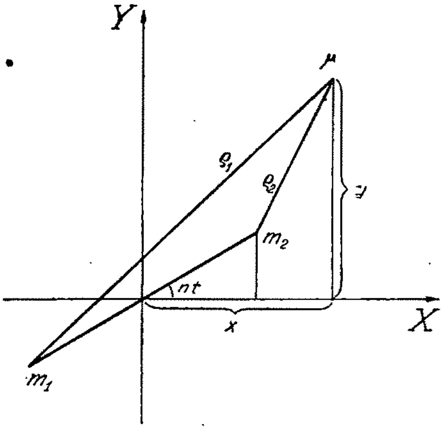


Fig. 9.

It now turned out that the study of the class of orbits shown in Fig. 3 had very extensive consequences. We see how an asymptotic-periodic orbit acts as the boundary orbit of a class of periodic orbits. The following idea was now rather obvious: If we can find other asymptotic orbits, which are also periodic (with infinitely long period), will it not then turn out that these orbits are boundary orbits of classes of periodic orbits?

The way that had to be followed now was very simple. We computed various orbits, that were asymptotic from the point of libration L_4.

We started the calculations with the aid of the theoretical expressions for the motion infinitely close to the libration point.

Fig. 9 defines the problem in the fixed coördinate system with center of gravity of m_1 and m_2 at the origin. The differential equations of the motion of the infinitely small mass are, in the fixed coördinate system, and with our choice of units, as follows:

$$(1)\quad \begin{cases} \dfrac{d^2x}{dt^2} = -4\left\{\dfrac{x+\cos t}{\rho_1^3} + \dfrac{x-\cos t}{\rho_2^3}\right\} \\ \dfrac{d^2y}{dt^2} = -4\left\{\dfrac{y+\sin t}{\rho_1^3} + \dfrac{y-\sin t}{\rho_2^3}\right\} \\ \rho_1^2 = (x+\cos t)^2 + (y+\sin t)^2 \\ \rho_2^2 = (x-\cos t)^2 + (y-\sin t)^2. \end{cases}$$

Referred to the coördinate system rotating with the line m_1m_2 the equations are:

$$(2)\quad \begin{cases} \dfrac{d^2\xi}{dt^2} - 2\dfrac{d\eta}{dt} - \xi = -4\left\{\dfrac{\xi+1}{\rho_1^3} + \dfrac{\xi-1}{\rho_2^3}\right\} \\ \dfrac{d^2\eta}{dt^2} + 2\dfrac{d\xi}{dt} - \eta = -4\left\{\dfrac{\eta}{\rho_1^3} + \dfrac{\eta}{\rho_2^3}\right\}. \\ \rho_1^2 = (\xi+1)^2 + \eta^2 \\ \rho_2^2 = (\xi-1)^2 + \eta^2. \end{cases}$$

If we transfer the origin of the coördinate system to L_4 and neglect powers of the coördinates higher than the first we obtain the following equations of motion:

$$(3)\quad \begin{cases} \dfrac{d^2x}{dt^2} - 2\dfrac{dy}{dt} = \dfrac{3}{4}x \\ \dfrac{d^2y}{dt^2} + 2\dfrac{dx}{dt} = \dfrac{9}{4}y. \end{cases}$$

which are, therefore, the differential equations of the infinitesimal motion near L_4. The solution for the ingoing motion is given in terms of the functions

$$e^{-at}\sin bt \quad\text{and}\quad e^{-at}\cos bt,$$

that for the outgoing motion in terms of

$$e^{+at}\sin bt \quad\text{and}\quad e^{+at}\cos bt,$$

i. e. we have an asymptotic motion.

We follow the motion with the aid of the infinitesimal equations up to a point, which is very close to L_4, and then continue the solution with the aid of numerical integration of the exact differential equations (2).

If we follow the motion until the orbit cuts the X-axis for the first time, we shall, in case the orbit should be perpendicular to the X-axis, have a periodic orbit–due to the symmetry of the differential equations. The orbit will continue symmetrically with regard to the X-axis, and asymptotically approach L_5. We may imagine the orbit continued in an orbit symmetrical with respect to the Y-axis, back to L_4, where the motion started.

The problem of finding such orbits, which are perpendicular to the X-axis, presents no difficulties. The method is to compute a number of orbits that cut the X-axis under various angles, interpolating finally between such of these orbits as make an angle of 90° with the X-axis.

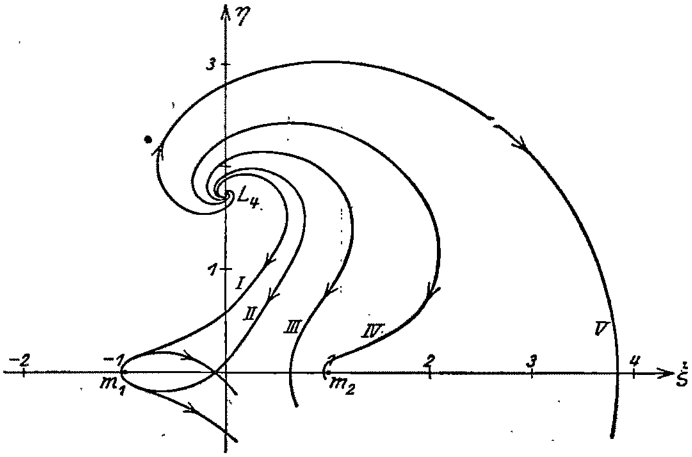


Fig. 10.

The 5 asymptotic-periodic orbits which are already perpendicular to the x-axis after a quarter of a revolution.

We have in Copenhagen made a large number of such calculations. Plate II gives the essential of the results obtained. Examining this drawing we find that 5 orbits exist, not more and not less, which are already perpendicular to the X-axis after a quarter of a revolution. Fig. 10 shows these five asymptotic-periodic orbits (for the sake of simplicity only about half of the orbits have been reproduced).

The question now arises, what does it mean that, in this way, a number of orbits exist, that are periodic (with infinitely long periods) and at the same time asymptotic to and from L_4 and L_5? Now, if for instance we investigate

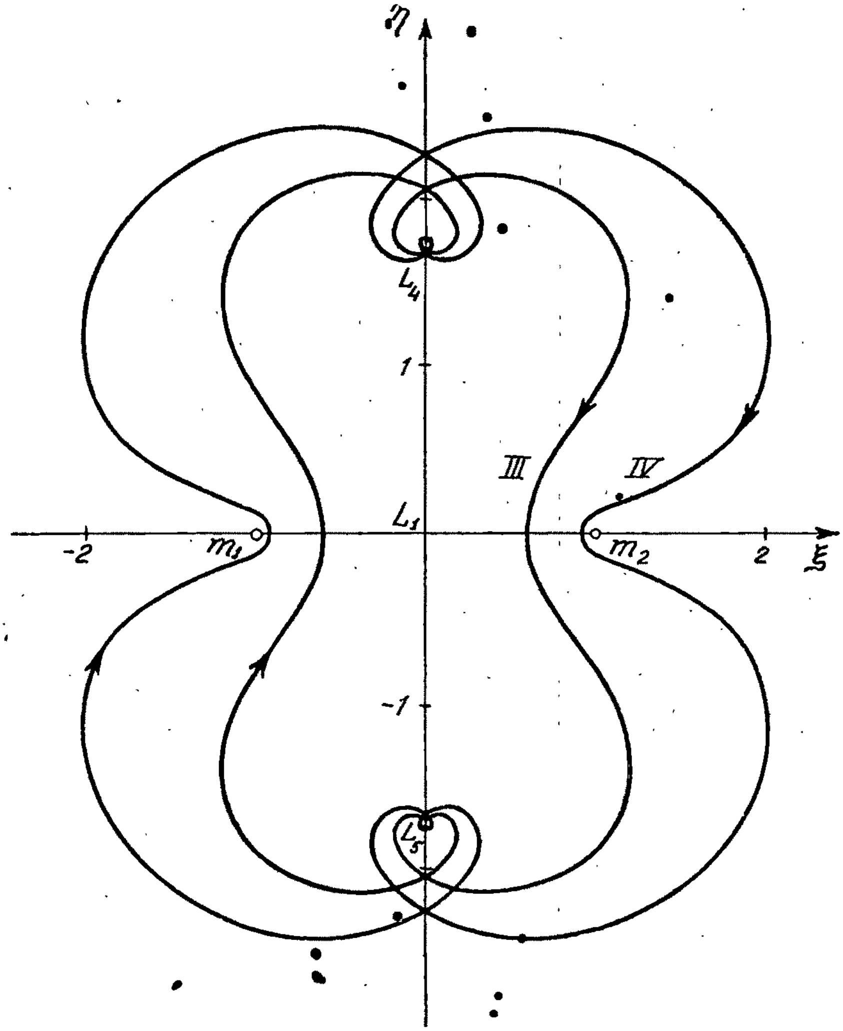


Fig. 11.

Two asymptotic orbits as boundary orbits of a class of periodic orbits.

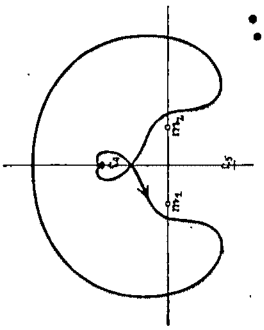


Fig. 1.

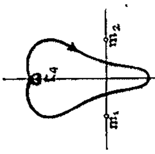


Fig. 2.

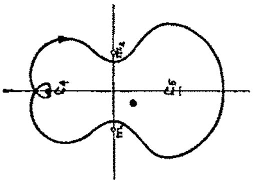


Fig. 3.

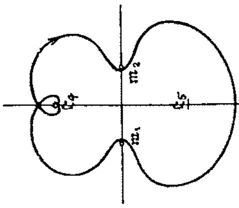


Fig. 4.

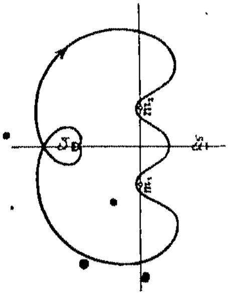


Fig. 5.

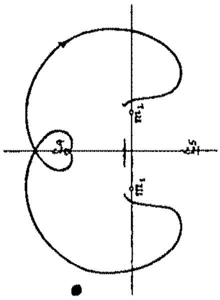


Fig. 6.

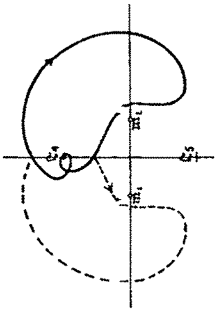


Fig. 7-8.

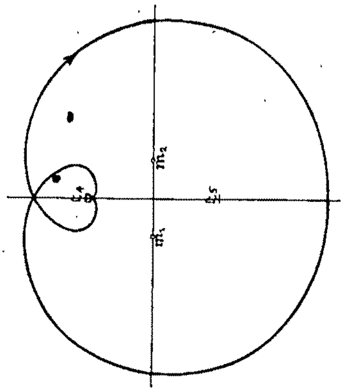


Fig. 9.

PLATE III.

orbit V in Fig. 10, we find that this orbit is identical with the boundary orbit that was found during the study of the class of orbits mentioned before, which had its origin at infinity.

A more detailed investigation showed that the four other orbits are also boundary orbits of classes of periodic orbits. Thus we have for instance a class of periodic orbits the boundary orbits of which are the orbits III and IV in Fig. 10. Fig. 11 reproduces these two boundary orbits drawn completely. In between these two boundary orbits we have to imagine a class of orbits—infinite in number—the boundary orbits of which are the two asymptotic orbits.

Herewith we have been led to a quite new principle of finding classes of periodic orbits. Formerly the starting point was those periodic orbits, which were known to exist near the masses or the libration points, and through laborious numerical calculations the evolution of each separate class was followed, until the whole evolution of the class up to the natural boundary curves had been covered. Now we are faced with the inverse problem: Once we have found an asymptotic orbit, which is also periodic (with infinitely long period), then we also have a boundary orbit for a whole class of periodic orbits.

We shall not enter into further details, but only touch upon one further point. We found the 5 boundary orbits in Fig. 10 by looking for asymptotic orbits which cut the X-axis at right angles. Of course we can also investigate the problem of finding orbits asymptotic to L_4, which do not cut the X-axis at right angles, but do intersect the Y-axis perpendicularly, one quadrant later. We shall have to deal, in that case, with orbits that are not—like those 5 mentioned before—asymptotic to L_5, but which, having cut the Y-axis, continue in a branch symmetrical to the first with respect to the Y-axis, and end in the libration point L_4, where they started. Our investigations now showed that 6 such orbits exist. They can be found on Plate II, and they are all given together on Plate III (No. 2, 3, 4, 5, 6, and 9). All of these orbits are boundary orbits of classes of periodic orbits, and of course the reasoning can be extended: We can look for all orbits, asymptotic out of L_4, which on cutting the X-axis for the second time are perpendicular to it, and orbits, which on cutting the Y-axis for the second time are at right angles to it, etc., to infinity. In this way we will get boundary orbits of an infinitely increasing number of classes of more and more complicated periodic orbits.

Hill's investigations of the variational curve of the moon and of the problem of the curves of zero velocity introduced new principles into lunar theory. Poincaré showed the importance of periodic orbits and proved the existence of asymptotic orbits in the problem of three bodies. Thiele-Burrau and G. H. Darwin introduced the methods of numerical integration in the investigations of the restricted problem. In these facts we have the origin of the work on the restricted problem that has been carried out at the Copenhagen Observatory.

ASTRONOMISK OBSERVATORIUM,
KØBENHAVN.

ZUR THEORIE DER DIRICHLETSCHEN REIHEN.

Von Otto Toeplitz.

Die mathematischen Gedanken, die G. W. Hill in seiner berühmten Arbeit über das Mondperigäum entwickelt hat, sind eines der Einfallstore geworden, durch die man den Zugang zum Bereich der unendlich vielen Veränderlichen gefunden hat. Von dieser Arbeit Hills angeregt, hat H. Poincaré die allgemeinen Sätze über unendliche Determinanten entdeckt, die H. von Koch zu einer geschlossenen Theorie abgerundet hat und die I. Fredholm als ein Vorbild bei seiner Konzeption der allgemeinen Theorie der linearen Integralgleichungen dienten. Der Allgemeinheitsgrad dieser Sätze, der in dieser historischen Entwicklung als ein entscheidendes Moment erscheint, ist bei Hill explizite nicht anzutreffen; er handelt scheinbar von einem speziellen Gleichungssystem. Das ist nur scheinbar der Fall. In Wahrheit hat er als ein genialer numerischer Rechner aus den Zahlen seines Spezialfalles die allgemeinen Gedanken herausschauen lassen, die er an der Hand des Sonderfalls erkannte, und vielleicht sind diese Gedanken von der Theorie der unendlich vielen Variablen noch heute nicht voll ausgeschöpft worden.

Obgleich die vorliegende Arbeit zu der oben berührten Aufgabe keinen direkten Bezug hat, soll sie doch Zeugnis ablegen von dem Andenken, das die Lehre von den unendlich vielen Variablen dem Geiste Hills schuldet.

Die Arbeit ist der unveränderte Abdruck von einem Teil eines Manuskripts, das im Jahre 1918 niedergeschrieben worden ist, in den Jahren 1918-1922 einer Reihe von Fachgenossen vorlag und im mathematischen Seminar der Hamburger Universität referiert worden ist. Herr A. Wintner, dem ich für die Anregung zum Abdruck und für die Besorgung der Redaktion besonderen Dank schulde, hat dieses Manuskript im September 1929 kennen gelernt.

1. Die *D*-Formen. Es sei $f(ti)$ eine für $-\infty < t < +\infty$ erklärte Funktion, für welche alle Hadamardschen Koeffizienten[1]

$$c(\lambda) = \mathfrak{M}\{f(ti)e^{i\lambda t}\} \qquad (-\infty < \lambda < +\infty)$$

existieren, wobei $\mathfrak{M}$ die Operation der Mittelbildung bezeichnet:

$$\mathfrak{M}\{\cdot\ \cdot\ \cdot\} = \lim_{\omega=+\infty} \frac{1}{2\omega}\int_{-\omega}^{\omega} \cdot\ \cdot\ \cdot\ dt.$$

Und zwar soll $f(ti)$ derart sein, dass die Hadamardschen Koeffizienten $c(\lambda)$ ausser für $\lambda = \lg 1, \lg 2, \lg 3, \cdots$ verschwinden. Der Koeffizient $c(\lg n)$

---

[1] J. Hadamard, *Acta Mathematica*, Bd. 22 (1899), insb. S. 60-63.

werde kurz als c_n bezeichnet, also insbesondere $c(0)$ als c_1. Speziell wird also nach Hadamard und Schnee eine solche Belegung stets dann vorliegen, wenn man eine spezielle Dirichletsche Reihe $\Sigma c_n/n^s$ hat, die auf $s = ti$ absolut oder über $s = ti$ hinaus bedingt konvergiert, und wenn man ihren Summenwert $f(ti)$ als Belegung nimmt.

Alsdann betrachte man den nicht-negativen Ausdruck[2]

$$
\begin{aligned}
(1)\quad & (\bar{x}_1 + 2^{ti}\bar{x}_2 + 3^{ti}\bar{x}_3 + \cdots + n^{ti}\bar{x}_n)(x_1 + 2^{-ti}x_2 + 3^{-ti}x_3 + \cdots + n^{-ti}x_n) \\
&= (x_1\bar{x}_1 + \cdots + x_n\bar{x}_n) + (x_1\bar{x}_2 + x_2\bar{x}_4 + x_3\bar{x}_6 + \cdots)2^{ti} \\
&\quad + (x_2\bar{x}_3 + x_4\bar{x}_6 + \cdots)(3/2)^{ti} + \cdots + x_1\bar{x}_n n^{ti} \\
&\quad + (\bar{x}_1 x_2 + \bar{x}_2 x_4 + \bar{x}_3 x_6 + \cdots)2^{-ti} \\
&\quad + (\bar{x}_2 x_3 + \bar{x}_4 x_6 + \cdots)(3/2)^{-ti} + \cdots + \bar{x}_1 x_n n^{-ti},
\end{aligned}
$$

multipliziere ihn mit $f(ti)$ und bilde den Integralmittelwert nach t:

$$
(2)\quad D_n(x, \bar{x}) = \mathfrak{M}\{f(ti)(\bar{x}_1 + 2^{ti}\bar{x}_2 + \cdots + n^{ti}\bar{x}_n)(x_1 + 2^{-ti}x_2 + \cdots + n^{-ti}x_n)\}.
$$

Nun ist vorausgesetzt, dass

$$
\mathfrak{M}(f(ti)n^{ti}) = c_n \text{ oder } = 0,
$$

je nachdem n positiv-ganz ist oder nicht. Führt man also in (2) unter dem Integralzeichen statt der linken Seite von (1) die rechte Seite ein, so verschwindet die Mehrzahl der Summanden, und es bleibt:

$$
\begin{aligned}
(3)\quad D_n(x, \bar{x}) = c_1(x_1\bar{x}_1 + \cdots + x_n\bar{x}_n) &+ c_2(x_1\bar{x}_2 + x_2\bar{x}_4 + x_3\bar{x}_6 + \cdots) \\
&+ c_3(x_1\bar{x}_3 + x_2\bar{x}_6 + \cdots) + \cdots + c_n(x_1\bar{x}_n).
\end{aligned}
$$

Ist nun überdies die Belegung $f(ti)$ beschränkt, $|f(ti)| \leqq M$, so kann man auf das Integral in (2) den ersten Mittelwertsatz der Integralgleichung anwenden, da der Faktor von $f(ti)$ eine reelle, nicht-negative Funktion ist, und erhält

$$
|D_n(x, \bar{x})| \leqq M\mathfrak{M}\{(\bar{x}_1 + 2^{ti}\bar{x}_2 + \cdots + n^{ti}\bar{x}_n)(x_1 + 2^{-ti}x_2 + \cdots + n^{-ti}x_n)\},
$$

also unter Benutzung der Relation (1):

$$
|D_n(x, \bar{x})| \leqq M(x_1\bar{x}_1 + \cdots + x_n\bar{x}_n),
$$

da $\mathfrak{M}\{e^{i\lambda t}\}$ verschwindet oder gleich eins wird, je nachdem $\lambda \neq 0$ oder $\lambda = 0$ ist.

Theorem I. *Befriedigt eine Belegung ausser den eingangs dieses Paragraphen aufgeführten Voraussetzungen noch die weitere* $|f(ti)| \leqq M$, *und ist* H_n *das Maximum des Ausdrucks* (3) *unter der Nebenbedingung* $x_1\bar{x}_1 + \cdots + x_n\bar{x}_n = 1$, *so ist* $\mathrm{H}_n \leqq M$ *für jedes* n.

---

[2] Hierbei sind die Summen in den Klammern alle soweit zu erstrecken, als die Indizes $\leqq n$ bleiben.

Bemerkung. Bedient man sich der Sprechweise der Theorie der unendlich vielen Variablen, so ist $D_n(x, \bar{x})$ nichts anderes als der n-te Abschnitt der Bilinearform von unendlich vielen Variablen

$$\begin{aligned} D(x, y) = c_1(x_1y_1 + x_2y_2 + \cdots) &+ c_2(x_1y_2 + x_2y_4 + \cdots) \\ &+ c_3(x_1y_3 + x_2y_6 + \cdots) + \cdots, \end{aligned}$$

nur dass in den n-ten Abschnitt $D_n(x, y)$ hiervon noch $y_1 = \bar{x}_1, \cdots, y_n = \bar{x}_n$ gesetzt ist. Nun besagt ein einfacher Hilfssatz[3] über Bilinearformen von endlich vielen Veränderlichen $A(x, y)$, dass das Maximum von $|A(x, y)|$ für $\Sigma x_a \bar{x}_a = 1$, $\Sigma y_a \bar{y}_a = 1$ höchstens gleich dem Doppelten des Maximums von $A(x, \bar{x})$ für $\Sigma x_a \bar{x}_a = 1$ ist; also liegt im vorliegenden Falle $|D_n(x, y)|$ unter $2M$, und Theorem I kann auch so formuliert werden:

Befriedigt eine Belegung die eingangs dieses Paragraphen aufgeführten Voraussetzungen, so soll die aus ihren Hadamardschen Koeffizienten c_n aufgebaute Bilinearform $D(x, y)$ als die zu der Belegung gehörige „*D-Form*" bezeichnet werden. *Ist die Belegung beschränkt,* $|f(ti)| \leqq M$, *so ist die D-Form im Hilbertschen Sinne beschränkt, und ihre obere Schranke liegt unter* $2M$.

Die Koeffizientenmatrix einer solchen D-Form hat die Eigentümlichkeit, dass die Stellen, die den nämlichen Koeffizienten tragen, je eine gerade Linie erfüllen und dass alle diese geraden Linien ein Strahlenbüschel bilden mit endlichem Zentrum:[4]

$$D = \left\| \begin{array}{ccccccccccccc} c_1 & c_2 & c_3 & c_4 & c_5 & c_6 & c_7 & c_8 & c_9 & c_{10} & c_{11} & c_{12} & \cdots \\ 0 & c_1 & 0 & c_2 & 0 & c_3 & 0 & c_4 & 0 & c_5 & 0 & c_6 & \cdots \\ 0 & 0 & c_1 & 0 & 0 & c_2 & 0 & 0 & c_3 & 0 & 0 & c_4 & \cdots \\ 0 & 0 & 0 & c_1 & 0 & 0 & 0 & c_2 & 0 & 0 & 0 & c_3 & \cdots \\ 0 & 0 & 0 & 0 & c_1 & 0 & 0 & 0 & 0 & c_2 & 0 & 0 & \cdots \\ 0 & 0 & 0 & 0 & 0 & c_1 & 0 & 0 & 0 & 0 & 0 & c_2 & \cdots \\ 0 & 0 & 0 & 0 & 0 & 0 & c_1 & 0 & 0 & 0 & 0 & 0 & \cdots \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & c_1 & 0 & 0 & 0 & 0 & \cdots \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & c_1 & 0 & 0 & 0 & \cdots \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & c_1 & 0 & 0 & \cdots \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & c_1 & 0 & \cdots \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & c_1 & \cdots \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \end{array} \right\|$$

---

[3] Vgl. eine inzwischen in der *Mathematischen Zeitschrift* (Bd. 2 (1918), S. 187) erschienene Arbeit.

[4] Während sie bei den L-Formen Parallelstrahlenbüschel bilden.

Das System dieser D-Formen oder D-Matrizen ist isomorph den rein formal betrachteten zugehörigen Dirichletschen Reihen, d. h. der Summe und dem formalen Produkt zweier spezieller Dirichletscher Reihen gehört die Summe bzw. das Produkt der zugehörigen D-Formen (im Sinne des Matrizenkalküls) zu.

Nach einem bekannten Hilfssatze von Weierstrass kann man aus der Positivität des Faktors von $f(ti)$ in (2) mehr entnehmen:

Zusatz. *Ist $\mathfrak{R}$ der kleinste konvexe Bereich, der alle Werte der Belegung $f(ti)$ umschliesst, so liegt jeder Wert, den $D_n(x, \bar{x})$ für $\Sigma x_a \bar{x}_a = 1$ annimmt, in $\mathfrak{R}$.*

2. Umkehrung von Theorem I für absolut konvergente Dirichletsche Reihen. Ist $f(s) = \Sigma c_n/n^s$ absolut konvergent für $s = ti$, ist m. a. W. $\Sigma \mid c_n \mid$ konvergent, so ist $\Sigma c_n/n^s$ a fortiori gleichmässig konvergent, und es folgt aus den Ergebnissen von Hadamard, dass $f(ti)$ alle Voraussetzungen von § 1, auch die der Beschränktheit, erfüllt. Es gilt dann:

Theorem II. *Ist $\Sigma \mid c_n \mid$ konvergent und wird $\Sigma c_n/n^{ti}$ gesetzt, so kommt der in Theorem I mit $D_n(x, \bar{x})$ bezeichnete Ausdruck unter der Nebenbedingung $\Sigma x_a \bar{x}_a = 1$ und bei hinreichend grossem n jedem Wert der Belegung $f(ti)$ beliebig nahe.*

Der Beweis werde mit dem naheliegenden Versuch begonnen

(1) $\qquad x_1 = 1/\sqrt{\alpha}, \cdots, x_a = 1/\sqrt{\alpha}$, alle weiteren $x_n = 0$.

Dann ist, wenn $n \geqq \alpha$, $x_1\bar{x}_1 + \cdots + x_n\bar{x}_n = 1$. Ferner kommt in $D_n(x, \bar{x})$ offenbar c_1 genau α mal vor, jedesmal mit dem Faktor $1/\sqrt{\alpha} \cdot 1/\sqrt{\alpha} = 1/\alpha$, das gibt zusammen c_1, während c_2 genau $\frac{1}{2}\alpha$ mal vorkommt, wenn α gerade ist, und $\frac{1}{2}(\alpha - 1)$ mal bei ungeradem α, u. s. f.; also ist für das Wertsystem (1)

$$D_n(x, \bar{x}) = c_1 + \tfrac{1}{2}c_2 + \tfrac{1}{3}c_3 + \cdots + \frac{1}{\alpha}c_a + \Delta_a,$$

wo die kleine Abweichung Δ_a jedenfalls der Abschätzung unterliegt

$$\mid \Delta_a \mid \leqq \mid c_2 \mid \frac{1}{2\alpha} + \mid c_3 \mid \tfrac{2}{3} \cdot \frac{1}{\alpha} + \cdots + \mid c_a \mid \frac{1}{\alpha} \leqq \frac{\mid c_2 \mid + \cdots + \mid c_a \mid}{\alpha}.$$

Wegen der Konvergenz von $\Sigma \mid c_a \mid$ ist $\lim\limits_{a=\infty} \mid c_a \mid = 0$, also nach dem Satz vom arithmetischen Mittel auch $\lim\limits_{a=\infty} \Delta_a = 0$, und $D_n(x, \bar{x})$ nähert sich für das Wertsystem (1) und mit wachsendem α und $n \geqq \alpha$ immer mehr dem Werte

(2) $$c_1 + \tfrac{1}{2}c_2 + \tfrac{1}{3}c_3 + \cdots = f(1).$$

So könnte man von $f(1 + ti)$, wie eben für $f(1)$, dasjenige einsehen, was von $f(0)$ und $f(0 + ti)$ gezeigt werden soll.

Es soll deshalb der Versuch mit anderen, komplizierteren Wertsystemen x_n

wiederholt werden, die übrigens den Vorteil haben werden, dass die lästige Abschätzung eines Δ_α nicht auftritt. Es sei zuerst, mit $|\eta| = 1$,

$$(3)\qquad x_1 = 1/\sqrt{\alpha},\ x_2 = \eta/\sqrt{\alpha},\ x_4 = \eta^2/\sqrt{\alpha}, \cdots, x_{2^{\alpha-1}} = \eta^{\alpha-1}/\sqrt{\alpha},$$
$$\text{alle anderen } x_\nu = 0.$$

So erhält man neben $\sum_{\nu=1}^{n} x_\nu \bar{x}_\nu = 1$, wenn n hinreichend gross ist, $n \geqq 2^\alpha$,

$$(4)\qquad D_n(x, \bar{x}) = c_1 + \frac{\alpha-1}{\alpha}\bar{\eta}c_2 + \frac{\alpha-2}{\alpha}\bar{\eta}^2 c_4 + \cdots + \frac{1}{\alpha}\bar{\eta}^{\alpha-1}c_{2^{\alpha-1}},$$

oder, wenn $\eta = 2^{ti}$ gesetzt wird,

$$(5)\quad D_n(x, \bar{x}) = c_1 + \frac{\alpha-1}{\alpha}\,\frac{c_2}{2^{ti}} + \frac{\alpha-2}{\alpha}\,\frac{c_4}{4^{ti}} + \cdots + \frac{1}{\alpha}\,c_{2^{\alpha-1}}\,\frac{1}{2^{(\alpha-1)ti}},$$

d. h. das α-te arithmetische Mittel der unendlichen Reihe

$$(6)\qquad c_1 + \frac{c_2}{2^{ti}} + \frac{c_4}{4^{ti}} + \frac{c_8}{8^{ti}} + \cdots.$$

Da diese eine Teilreihe der als absolut konvergent vorausgesetzten Reihe $\Sigma c_n/n^{ti}$ ist, ist sie auch konvergent, und somit auch die Folge ihrer arithmetischen Mittel, und diese konvergiert gegen denselben Wert. Dem Wert dieser Teilreihe kommt aber zunächst die Form $D_n(x, \bar{x})$ im angegebenen Sinne beliebig nahe.

Es sei nun weiter bei hinreichend hohem n ($n \geqq 2^\alpha \cdot 3^\beta$):

$$|\eta| = 1, \qquad |\theta| = 1$$

und

$$(7)\quad \begin{array}{llll}
x_1 = \dfrac{1}{\sqrt{\alpha\beta}}, & x_2 = \dfrac{\eta}{\sqrt{\alpha\beta}}, & x_4 = \dfrac{\eta^2}{\sqrt{\alpha\beta}}, \cdots, & x_{2^{\alpha-1}} = \dfrac{\eta^{\alpha-1}}{\sqrt{\alpha\beta}}, \\
x_3 = \dfrac{\theta}{\sqrt{\alpha\beta}}, & x_{3.2} = \dfrac{\theta\eta}{\sqrt{\alpha\beta}}, & x_{3.4} = \dfrac{\theta\eta^2}{\sqrt{\alpha\beta}}, \cdots, & x_{3.2^{\alpha-1}} = \dfrac{\theta\eta^{\alpha-1}}{\sqrt{\alpha\beta}}, \\
\cdot\ \cdot\ \cdot & \cdot\ \cdot\ \cdot & \cdot\ \cdot\ \cdot & \cdot\ \cdot\ \cdot \\
x_{3^{\beta-1}} = \dfrac{\theta^{\beta-1}}{\sqrt{\alpha\beta}}, & x_{3^{\beta-1}.2} = \dfrac{\theta^{\beta-1}\eta}{\sqrt{\alpha\beta}}, & x_{3^{\beta-1}.4} = \dfrac{\theta^{\beta-1}\eta^2}{\sqrt{\alpha\beta}}, \cdots; & x_{3^{\beta-1}.2^{\alpha-1}} = \dfrac{\theta^{\beta-1}\eta^{\alpha-1}}{\sqrt{\alpha\beta}},
\end{array}$$

alle anderen $x_\nu = 0$, so ist wiederum $\Sigma x_\nu \bar{x}_\nu = 1$, und $D_n(x, \bar{x})$ wird gleich

$$(8)\quad \begin{aligned}
& c_1 + c_2\bar{\eta}\frac{(\alpha-1)\beta}{\alpha\beta} + c_4\bar{\eta}^2\frac{(\alpha-2)\beta}{\alpha\beta} + \cdots + c_{2^{\alpha-1}}\bar{\eta}^{\alpha-1}\frac{\beta}{\alpha\beta} \\
& + c_3\bar{\theta}\frac{\alpha(\beta-1)}{\alpha\beta} + c_{3.2}\bar{\theta}\bar{\eta}\frac{(\alpha-1)(\beta-1)}{\alpha\beta} + c_{3.4}\bar{\theta}\bar{\eta}^2\frac{(\alpha-2)(\beta-1)}{\alpha\beta} \\
& \qquad + \cdots + c_{3.2^{\alpha-1}}\bar{\theta}\bar{\eta}^{\alpha-1}\frac{\beta-1}{\alpha\beta} \\
& + \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot \\
& + c_{3^{\beta-1}}\bar{\theta}^{\beta-1}\frac{\alpha}{\alpha\beta} + c_{3^{\beta-1}.2}\bar{\theta}^{\beta-1}\bar{\eta}\frac{\alpha-1}{\alpha\beta} + c_{3^{\beta-1}.4}\bar{\theta}^{\beta-1}\bar{\eta}^2\frac{\alpha-2}{\alpha\beta} \\
& \qquad + \cdots + c_{3^{\beta-1}.2^{\alpha-1}}\bar{\theta}^{\beta-1}\bar{\eta}^{\alpha-1}\frac{1}{\alpha\beta}.
\end{aligned}$$

Nun ist die Doppelreihe

$$
(9)\qquad
\begin{array}{l}
\, c_1 + c_2/2^{ti} + c_4/4^{ti} + \cdots \\
+ c_3 + c_6/6^{ti} + c_{12}/12^{ti} + \cdots \\
+ c_9 + c_{18}/18^{ti} + c_{36}/36^{ti} + \cdots \\
+ \cdot\;\cdot\;\cdot\;\cdot\;\cdot\;\cdot\;\cdot\;\cdot\;\cdot\;\cdot\;\cdot
\end{array}
$$

absolut konvergent als Teilreihe der absolut konvergenten Reihe $\Sigma c_n/n^{ti}$. Setzt man also $\theta = 3^{ti}$, $\eta = 2^{ti}$, so kann man zuerst bei festem β den Index α so gross wählen, dass (8) beliebig wenig abweicht von

$$
\begin{array}{l}
\quad\{c_1 + c_2(1/2^{ti}) + c_4(1/4^{ti}) + \cdots\} \\
+ [(\beta - 1)/\beta]\{c_3(1/3^{ti}) + c_{3.2}(1/6^{ti}) + c_{3.4}(1/12^{ti}) + \cdots\} \\
+ \cdot\;\cdot\;\cdot\;\cdot\;\cdot\;\cdot\;\cdot\;\cdot\;\cdot\;\cdot\;\cdot\;\cdot\;\cdot\;\cdot\;\cdot\;\cdot\;\cdot\;\cdot\;\cdot\;\cdot \\
+ (1/\beta)\{c_{3^{\beta-1}}[1/(3^{\beta-1})^{ti}] + c_{3^{\beta-1}.2}[1/(3^{\beta-1}.2)^{ti}] + c_{3^{\beta-1}.4}[1/(3^{\beta-1}.4)^{ti}] + \cdots\},
\end{array}
$$

d. h. von dem β-ten arithmetischen Mittel aus den Zeilensummen der Doppelreihe (9). Endlich kann man also β so gross werden lassen, dass man den Summenwert der gesamten Doppelreihe (9) beliebig nahe kommt. Der Wert von (9) wird somit durch Werte von $D_n(x, \bar{x})$ bei hinreichend hohem n approximiert; (9) ist aber schon eine umfassendere Teilreihe von $\Sigma c_n/n^{ti}$ als (6).

Setzt man allgemeiner, indem man mit p_ν die ν-te Primzahl bezeichnet, $|\eta_k| = 1$ für $k = 1, \cdots, \nu$,

$$
x_{p_1^{u_1}\ldots p_\nu^{u_\nu}} = \eta_1^{u_1}\cdots\eta_\nu^{u_\nu}/\sqrt{\alpha_1\cdots\alpha_\nu} \quad \text{für}\ 0 \leqq u_1 < \alpha_1, \cdots, 0 \leqq u_\nu < \alpha_\nu,
$$

alle anderen $x_k = 0$, so folgt durch eine analoge Schlussweise, dass man mit $D_n(x, \bar{x})$ bei hinreichend hohem n diejenige Teilreihe von $\sum\limits_{n=1}^{\infty} c_n/n^{ti}$ approximiert, deren Glieder einen nur durch $p_1, \cdots, p_\nu$ teilbaren Index n haben. Nun ist die ganze Reihe absolut konvergent; man kann also N so gross wählen, dass $\sum\limits_{n=N+1}^{\infty} | c_n/n^{ti} | \leqq \frac{1}{2}\epsilon$ ist. Alsdann wähle man so gross, dass alle in den Zahlen $1, 2, 3, \cdots, N$ aufgehenden Primfaktoren unter $p_1, \cdots, p_\nu$ enthalten sind. Dann erhält man als Wert von $D(x, \bar{x})$ eine Teilreihe, die zunächst *alle* Glieder von $\sum\limits_{n=1}^{N} c_n/n^{ti}$ enthält und dann noch manche von dem Rest; die letzteren zusammen aber betragen weniger als $\frac{1}{2}\epsilon$; also ist $\sum\limits_{n=1}^{\infty} c_n/n^{ti}$ von einem Wert von $D_n(x, \bar{x})$ um weniger als ϵ entfernt, q. e. d.

Die Grössen $H_1, H_2, \cdots$ bilden eine Folge zunehmender, positiver Zahlen; nach Theorem I liegen sie alle unter M; also haben sie einen Grenzwert

$\lim\limits_{n=\infty} \mathrm{H}_n = \mathrm{H} \leqq M$. Theorem II gestattet nun, wenn M die genaue obere Grenze von $|f(ti)|$ ist, zu folgern, dass $\mathrm{H} = M$ ist:

FOLGERUNG. *Ist M die obere Grenze des Betrages von $\Sigma c_n/n^{ti} = f(ti)$ und $\lim\limits_{n=\infty} \mathrm{H}_n = \mathrm{H}$, so ist $\mathrm{H} = M$.*

3. **Das Vertikalenmaximum.** Die bisher aufgestellten Theoreme gestatten bereits einige Anwendungen auf die Theorie der Dirichletschen Reihen.

DEFINITION. Ich sage, eine Belegung $\phi(t)$, $-\infty < t < +\infty$, nimmt einen Wert „asymptotisch" an, wenn sie ihm ausserhalb jedes Intervalls $-T \leqq t \leqq +T$ beliebig nahe kommt. Im gleichen Sinne spreche ich von dem „asymptotischen Maximum" von $|\phi(t)|$, von dem „asymptotischen Wertvorrat" u. s. w.

Auf Grund dieser Terminologie kann man eine Verschärfung, deren Theorem I fähig ist, folgendermassen formulieren:

THEOREM III. *Ist $f(ti)$ eine Belegung, die den Voraussetzungen am Eingang von § 1 genügt, ist $\tilde{M}$ das asymptotische Maximum von $|f(ti)|$ und $\mathrm{H} = \lim \mathrm{H}_n$ dieselbe Grösse wie in Theorem I, so ist $\mathrm{H} \leqq \tilde{M} \leqq M$.*

ZUSATZ. *Ist K der kleinste konvexe Bereich, der den asymptotischen Wertvorrat von $f(ti)$ enthält, so liegen alle Werte, die $D_n(x, \bar{x})$ für $\Sigma x_\alpha \bar{x}_\alpha = 1$ annimmt, in K (das ganz in $\mathfrak{K}$ enthalten ist).*

Beweis. Ist $\tilde{M}$ das asymptotische Maximum von $|f(ti)|$, so besagt dies: wenn $\epsilon > 0$ beliebig vorgegeben ist, kann man ein solches Intervall $-T \leqq t \leqq +T$ finden, dass die obere Grenze von $|f(ti)|$ ausserhalb dieses Intervalls $\leqq \tilde{M} + \epsilon$ ist. Betrachtet man nun neben der Belegung $f(ti)$ die andere $g(ti)$, die für $-T \leqq t \leqq +T$ den Wert 0 hat, ausserhalb dieses Intervalls aber mit $f(ti)$ übereinstimmt, so ist die obere Grenze von $|g(ti)| \leqq \tilde{M} + \epsilon$. Dann existieren aber die Hadamardschen Koeffizienten von $g(ti)$ und sind genau dieselben wie die von $f(ti)$; es wird also auch zu $g(ti)$ dieselbe D-Form gehören, wie zu $f(ti)$, mit derselben oberen Grenze H. Wendet man also Theorem I auf $g(ti)$ an, so erhält man $\mathrm{H} \leqq \tilde{M} + \epsilon$, und da ϵ beliebig klein gewählt werden konnte, $\mathrm{H} \leqq \tilde{M}$.

Nun ist unter den Voraussetzungen von Theorem II $\mathrm{H} = M$; nach Theorem III ist $\mathrm{H} \leqq \tilde{M}$, und seiner Definition nach ist $\tilde{M} \leqq M$; also ist auch $\mathrm{H} = \tilde{M}$:

SATZ 1. *Im Gebiete absoluter Konvergenz ist für eine Dirichletsche Reihe $f(s) = \Sigma c_n/n^s$ längs jeder Vertikalen $s = \sigma + ti$ die obere Grenze von*

$|f(\sigma+ti)|$ mit dem asymptotischen Maximum $\tilde{M}(\sigma)$ dieser Belegung identisch. Genauer liegen alle Werte von $f(\sigma+ti)$ innerhalb des kleinsten konvexen Bereiches $\mathrm{K}(\sigma)$, den man um den asymptotischen Wertvorrat von $f(\sigma+ti)$ legen kann.

Dieser Satz enthält als Folgerungen einige Theoreme, die Herr Bohr 1910 bewiesen hat. Für $\zeta(s)=\Sigma 1/n^s$ nämlich ist $\Re(s)>1$ das Gebiet absoluter Konvergenz, und da die Koeffizienten reell-positiv sind, ist $|\zeta(\sigma+ti)|\leqq\zeta(\sigma)$, also $\tilde{M}(\sigma)=\zeta(\sigma)$. Satz 1 zeigt nun, dass $\zeta(\sigma+ti)$ für beliebig hohe t dem Wert $\zeta(\sigma)$ beliebig nahe kommt. Da nun $\zeta(\sigma)$ unbegrenzt wächst, wenn σ von rechts an 1 heranrückt, ist also $\zeta(s)$ in der Halbebene $\Re(s)>1$ unbeschränkt. Entsprechendes gilt für jede Reihe $\Sigma c_n/n^s$ mit lauter nichtnegativen reellen Koeffizienten.

Satz 2. *Innerhalb des Gebietes absoluter Konvergenz ist das Vertikalenmaximum $M(\sigma)$ eine nirgends ansteigende Funktion der Vertikalenabscisse σ.*

Beweis. Das Vertikalenmaximum $M(\sigma)$ ist nach dem Ende von § 2 identisch mit der oberen Grenze $\mathrm{H}(\sigma)$ der D-Form mit den Koeffizienten c_n/n^σ:

$$\begin{aligned} D(x,\bar{x};\sigma) = c_1(x_1\bar{x}_1+x_2\bar{x}_2+\cdots)+\frac{c_2}{2^\sigma}(x_1\bar{x}_2+x_2\bar{x}_4+\cdots) \\ +\frac{c_3}{3^\sigma}(x_1\bar{x}_3+x_2\bar{x}_6+\cdots)+\cdots \end{aligned}$$

für $\Sigma x_a\bar{x}_a=1$. Nun geht diese Form durch die Substitution $x_n=n^{-\epsilon}\xi_n$ in $D(\xi,\bar{\xi};\sigma+2\epsilon)$ über, d. h. in die D-Form zu der um 2ϵ weiter nach rechts gelegenen Vertikalen. Es ist $\Sigma x_n\bar{x}_n=\Sigma n^{-2\epsilon}\xi_n\bar{\xi}_n$. Ist also $\Sigma\xi_n\bar{\xi}_n\leqq 1$, so ist erst recht $\Sigma x_n\bar{x}_n\leqq 1$, aber nicht umgekehrt. Ist also $|D(x,\bar{x};\sigma)|$ für $\Sigma x_n\bar{x}_n\leqq 1$ unter $M(\sigma)$ gelegen, so wird a fortiori $|D(\xi,\bar{\xi},\sigma+2\epsilon)|$ für $\Sigma\xi_n\bar{\xi}_n\leqq 1$ unter $M(\sigma)$ liegen, d. h. es ist $M(\sigma+2\epsilon)\leqq M(\sigma)$.

4. Umkehrung von Theorem I ohne einschränkende Annahme. Es sei eine D-Form $D(x,y)$ rein formal vorgelegt, und es sei lediglich von ihr vorausgesetzt, dass sie beschränkt ist. Daraus folgt an sich nach den allerersten Sätzen der Theorie der unendlich vielen Veränderlichen, dass die Quadratsumme $\Sigma|c_n|^2$ der Beträge ihrer Koeffizienten konvergiert; es sei aber keinerlei weitere Annahme, wie z. B. die der Konvergenz von $\Sigma|c_n|$ hinzugefügt. Alsdann kann man sich, anknüpfend an die Betrachtungen von § 2, derselben Methode bedienen, die Herr Bohr,[5] auf anderen Voraussetzungen fussend, zu einem ähnlichen Ziel benutzt hat.

Zuerst folgt aus der im § 1 im Anschluss an Theorem I gemachten Bemerkung, dass alle $| c_n | \leqq 2\mathrm{H}$ sind, wo H die obere Grenze von $| D(x, \bar{x}) |$ für $\Sigma x_a \bar{x}_a = 1$ ist. Sei nun, mit Herrn Bohr[5], zunächst rein formal:

$$f_1(s) = \sum_{n=1}^{\infty} c_{2^n} \frac{1}{2^{ns}}, \qquad f_2(s) = \sum_{n_1=1}^{\infty} \sum_{n_2=1}^{\infty} c_{2^{n_1}3^{n_2}} \frac{1}{2^{n_1 s} \cdot 3^{n_2 s}}, \cdots,$$

$$f_\nu(s) = \sum_{n_1=1}^{\infty} \cdots \sum_{n_\nu=1}^{\infty} c_{p_1^{n_1} \cdots p_\nu^{n_\nu}} \frac{1}{p_1^{n_1 s} \cdots p_\nu^{n_\nu s}}, \cdots,$$

so konvergiert also $f_1(s)$ für $\Re(s) > 0$ absolut und stellt daselbst eine analytische Funktion dar, ebenso $f_2(s), f_3(s), \cdots$.

Ferner folgt aus der schon erwähnten Konvergenz von $\Sigma | c_n |^2$ mit Hilfe der Schwarzschen Ungleichung, dass $\Sigma | c_n | / n^{\frac{1}{2}+\epsilon}$ für jedes $\epsilon > 0$ konvergiert, dass also $\Sigma c_n / n^s$ für $\Re(s) > \frac{1}{2}$ absolut konvergiert.

Endlich besagen (5) und (6) von § 2, dass das α-te arithmetische Mittel der Reihe $f_1(s)$ für $s = ti$ unter H gelegen ist; also ist $| f_1(s) | \leqq \mathrm{H}$ für $\Re(s) > 0$, und ebenso folgt aus § 2, dass $| f_\nu(s) | \leqq \mathrm{H}$ für $\Re(s) > 0$ und jedes $\nu = 1, 2, \cdots$.

Ist nun $1 + \tau i$ irgendein Punkt auf der Vertikalen mit der Abszisse 1, so betrachte man einen Kreis vom Radius $\frac{1}{2}$ um diesen Punkt. Innerhalb dieses Kreises ist $\Sigma c_n / n^s$ absolut konvergent; $f_1(s), f_2(s), \cdots$ sind Teilreihen davon; dieselben konvergieren also gleichmässig gegen $f(s)$. Nun besagt ein von Stieltjes stammender funktionentheoretischer Hilfssatz: wenn eine Folge analytischer Funktionen $f_1(s), f_2(s), \cdots$ in einem Kreise gleichmässig gegen $f(s)$ konvergiert und wenn dieselben Funktionen $f_\nu(s)$ in einem grösseren Gebiet dem Betrage nach unter einer gemeinsamen festen Grenze liegen, so konvergieren sie in jedem beschränkten abgeschossenen Teilbereich dieses grösseren Gebietes gleichmässig. Im vorliegenden Falle ergibt sich daraus, dass die obigen Funktionen $f_\nu(s)$ auch im Kreise vom Radius 1 um den Punkt $1 + \tau i$ gleichmässig konvergieren; die Grenzfunktion wird daselbst die analytische Fortsetzung von $f(s)$ sein und dem Betrage nach unter H liegen.

THEOREM IV. *Ist die zur Reihe* $\Sigma c_n / n^s$ *gehörige D-Form im Hilbertschen Sinne beschränkt und* H *die obere Grenze von* $D(x, \bar{x})$, *so konvergiert die Reihe selbst absolut für* $\Re(s) > \frac{1}{2}$ *und stellt eine analytische Funktion* $f(s)$ *dar, die bis an* $\Re(s) = 0$ *heran fortsetzbar ist und dem Betrage nach überall unter* H *liegt.*

BONN, GERMANY.

---

[5] H. Bohr, *Göttinger Nachrichten*, 1913.

MEAN MOTION.

By Hermann Weyl.

The problem of mean motion which arose from the theory of the secular perturbations of the planetary orbits [1] and which will here find its complete solution, deals with the *superposition of a finite number of epicycles* or of simple oscillations:

$$z = \sum_{k=1}^{n} a_k e(\vartheta_k), \tag{1}$$

$$\vartheta_k = \vartheta_k(t) = \lambda_k t + \alpha_k. \tag{2}$$

We use the abbreviation $e^{2\pi i\vartheta} = e(\vartheta)$. The *amplitudes* $a_k > 0$, *frequencies* λ_k, and *initial phases* α_k, are real constants; the *time* t is a real variable ranging from $-\infty$ to $+\infty$. The n-dimensional space of the real phases ϑ_k is to be construed as a *torus* Θ by identifying two points $(\vartheta) = (\vartheta_1, \cdots, \vartheta_n)$ and $(\vartheta') = (\vartheta'_1, \cdots, \vartheta'_n)$ which are congruent modulo 1:

$$\vartheta_k \equiv \vartheta'_k \; (\mathrm{mod}\, 1).$$

z is first a function of the n phases ϑ_k and then, after the substitution (2), a function of t. With the customary notation, polar coördinates are used in the z-plane:

$$z = r \cdot e(\phi).$$

Let us begin by discussing the phase function. (1) may be interpreted as depicting a (plane open) *linkage* $A_0A_1 \cdots A_n$ consisting of the links

$$A_{k-1}A_k = a_k e(\vartheta_k) \tag{3}$$

of fixed lengths a_k. $z = A_0A_n$ is the vector *spanned* by the linkage. Those points on the torus Θ for which $z = 0$ form what we call the *singular manifold*; it consists of all possible states of a *closed* n-linkage with the prescribed sides a_k. E being any area in the z-plane, the volume $v(E)$ of that part of the torus where $z(\vartheta_1 \cdots \vartheta_n)$ takes on values in E, indicates the phase probability of z lying in E. On account of central symmetry, it suffices to determine the probability

$$W(r) = W(r; a_1 \cdots a_n)$$

of the span z being in absolute value $< r$. By its very nature $W(r)$ is an increasing function of r varying between the limits 0 and 1. The strip $0 \leqq W \leqq 1$ in a (W, ϕ)-plane is a cylinder when points (W, ϕ), (W, ϕ') are regarded equal if $\phi \equiv \phi' \pmod 1$. The phase probability $v(E)$ is the area of the image of E on the (W, ϕ)-cylinder produced by the mapping

$$z = r \cdot e(\phi) \to (W(r), \phi).$$

A straightforward analysis by Fourier transforms, carried out by A. Wintner [2], results, for $n \geqq 2$, in the absolutely convergent expression

$$W(r; a_1 \cdots a_n) = r \int_0^\infty J_1(r\rho) \prod_{k=1}^n J_0(a_k\rho)\, d\rho \tag{4}$$

in terms of the Bessel functions J_0, J_1. The relationship to the random walk problem is obvious [3].

We now turn to the second standpoint with t as the independent variable, and assume that the λ_k satisfy no homogeneous linear relation

$$\sum_k m_k\lambda_k = 0 \tag{5}$$

with integral coefficients m_k. The *law of equidistribution* (Kronecker-Weyl) [4] then states that the straight line (2) fills the torus with uniform density. In other words, for any piece G of the torus with a Jordan volume G, the relative frequency with which the point $\vartheta(t)$, moving uniformly along a straight line, visits G, tends to G if the period of observation becomes infinitely large, or: *time-probability = phase-probability.* A slightly more general form of the proposition asserts that for any bounded Riemann integrable function $f(\vartheta_1 \cdots \vartheta_n)$

$$\frac{1}{t_2 - t_1} \int_{t_1}^{t_2} f((\vartheta(t)))\, dt \text{ tends to } \int_0^1 \cdots \int_0^1 f((\vartheta))\, d\vartheta_1 \cdots d\vartheta_n \tag{6}$$

with $t_2 - t_1 \to \infty$, or that the *time-average of f equals its phase-average.* Hence the considerations of the preceding paragraph determine the relative frequencies with which the "planet" $z(t)$ visits the various parts E of the z-plane.[1]

However, the problem of mean motion is slightly more intricate. To avoid ambiguities, let us suppose that the planet never passes through the origin (or that the straight line $\vartheta(t)$ does not hit the singular manifold). Then the continuous increment $\phi(t_2) - \phi(t_1)$ of the azimuth $\phi(t)$ of $z(t)$

---

[1] This is to be interpreted in a figurative sense, as in the astronomical application based on the Copernican rather than the Ptolomaic system $\phi(t)$ is not the longitude of the planet itself, but of its perihelion or its ascending node.

over any time interval $t_1 \leqq t \leqq t_2$ is uniquely determined. ϕ has a *mean motion* μ if

$$\frac{\phi(t_2)-\phi(t_1)}{t_2-t_1} \to \mu \text{ for } t_2-t_1 \to \infty.$$

We shall see that such a mean motion μ exists, depending linearly on the frequencies λ_k:

$$(7) \qquad \mu = W_1\lambda_1 + \cdots + W_n\lambda_n.$$

The complete statement is as follows.

THEOREM. *Assuming the frequencies λ_k to satisfy no linear relation* (5) *with integral coefficients m_k of sum* 0, *the azimuth ϕ of the superposition* (1), (2) *of epicycles has a mean motion*

$$(7) \qquad \mu = W_1\lambda_1 + \cdots + W_n\lambda_n.$$

The coefficients W_k depend on the amplitudes $a_1, \cdots, a_n$ only; they are $\geqq 0$ and their sum equals 1, *so that* (7) *is a certain average of the individual frequencies λ_k. More precisely, W_k is the probability*

$$(8) \qquad W_k = W(a_k; a_{k+1} \cdots a_n\, a_1 \cdots a_{k-1})$$

that an $(n-1)$-*linkage with the given sides $a_{k+1} \cdots a_n\, a_1 \cdots a_{k-1}$ spans a distance $< a_k$.*

For the proof we first return to our old hypothesis that the λ_k satisfy no linear homogeneous integral relation whatsoever, and then follow a method recently suggested by P. Hartmann, E. R. van Kampen, and A. Wintner [5]. We verify at once that the time derivative

$$\phi'(t) = \Re\left(\frac{1}{2\pi i}\frac{z'}{z}\right) \qquad \{\Re = \text{real part}\}$$

arises by the substitution (2) from the following phase function:

$$(9) \qquad \Phi((\vartheta)) = \Re\,\frac{\Sigma\lambda_k a_k e(\vartheta_k)}{\Sigma a_k e(\vartheta_k)}$$

which becomes dangerously infinite on the singular manifold. Were one allowed to apply to this function the principle of equidistribution (6) holding for bounded Riemann integrable functions f, one would at once obtain a mean motion

$$(10) \qquad \mu = \int_0^1 \cdots \int_0^1 \Phi((\vartheta))\, d\vartheta_1 \cdots d\vartheta_n.$$

The μ as defined by this formula depends linearly on the λ_k, (7), and its coefficients W_k are integrals of the following type:

$$W_n = \int_0^1 \cdots \int_0^1 \Re \left\{ \frac{a_n e(\vartheta_n)}{\sum_k a_k e(\vartheta_k)} \right\} d\vartheta_1 \cdots d\vartheta_n.$$

So far HKW. I shall now show how to carry out the integration by ϑ_n in W_n.

Indeed, for fixed $\vartheta_1, \cdots, \vartheta_{n-1}$ and variable $\vartheta_n = \vartheta$ we set

$$a_n = a, \qquad \sum_{k=1}^{n-1} a_k e(\vartheta_k) = b,$$

$$z = z(\vartheta) = b + a \cdot e(\vartheta).$$

The integrand is then

$$\Re\left(\frac{1}{2\pi i}\frac{z_\vartheta}{z}\right) \tag{11}$$

where the index ϑ designates derivation by ϑ. The integral of (11) by ϑ is $\int d\phi$ where ϕ denotes the direction of the radius vector from the origin O to the variable point $z(\vartheta)$ describing the circle of radius a around the center b, and hence we obtain 1 or 0 according as the origin lies within or without the circle. This distinction amounts to

$$|\, b \,| = |\, a_1 e(\vartheta_1) + \cdots + a_{n-1} e(\vartheta_{n-1}) \,| \lessgtr a_n. \tag{12}$$

We thus find

$$W_n = \int d\vartheta_1 \cdots d\vartheta_{n-1},$$

the integral extending over that part of the $(n-1)$-dimensional phase torus $(\vartheta_1, \cdots, \vartheta_{n-1})$ where

$$|\, a_1 e(\vartheta_1) + \cdots + a_{n-1} e(\vartheta_{n-1}) \,| < a_n,$$

or

$$W_n = W(a_n; a_1 \cdots a_{n-1}). \tag{13}$$

The formula (10) for $\Sigma W_k \lambda_k$ yields at once the equation

$$W_1 + \cdots + W_n = 1 \tag{14}$$

by taking $\lambda_1 = \cdots = \lambda_n = 1$. This result has its own interest independent of the problem of mean motion:

Given n positive lengths a_k, let W_k be the probability that an $(n-1)$-linkage with the sides $a_{k+1} \cdots a_n\, a_1 \cdots a_{k-1}$ spans a distance $< a_k$. Then the n probabilities $W_1, \cdots, W_n$ yield the sum 1.

Certain examples suggest that the equidistribution law (6) might not be applicable to functions with the type of singularity present in $\phi((\vartheta))$. HKW get around this difficulty by resorting to Birkhoff's general ergodic theorem. But then the formula is proved only with the exception of an unknown zero set in the space of the initial phases α_k. Our evaluation of the integral indicates, however, that the evil-spelling singularities of the integrand (9) are more or less bluff. For a complete proof of our theorem I therefore fall back on a much earlier and more elementary procedure inaugurated in 1909 by P. Bohl [6] for the lowest non-trivial case $n = 3$.

We slit the phase torus in the $(n-1)$-dimensional manifold of those points (ϑ) for which $z = z(\vartheta_1 \cdots \vartheta_n)$ is *real and negative.* On the slit torus the azimuth $\phi(\vartheta_1 \cdots \vartheta_n)$ is a continuous (periodic) single-valued function whose values lie between $-\frac{1}{2}$ and $+\frac{1}{2}$ (limits excluded). The boundary of the slit is the singular manifold. On the slit itself ϕ has a discontinuity, taking on the value $+\frac{1}{2}$ on the one, the $+$ side, and $-\frac{1}{2}$ on the $-$ side; it jumps by -1 or $+1$ according as one crosses the slit in the positive or negative sense (i. e. from the $+$ to the $-$ side, or in the opposite direction). Therefore $\phi(t_2) - \phi(t_1)$ differs by a term of absolute value $< \frac{1}{2}$ from the number of times $N(t_1t_2)$ the straight line $\vartheta(t)$ crosses the slit during the time interval $t_1 \leqq t \leqq t_2$. In computing $N(t_1t_2)$ the sense of crossing has to be taken into account. Let us erect a cylindrical trunk T over the slit as basis whose generator is the velocity vector $\lambda = (\lambda_1, \cdots, \lambda_n)$. Whenever our moving point $\vartheta(t)$ crosses the slit it will stay in T during the following unit of time, so that $N(t_1t_2)$ is essentially the duration of its stay in T between t_1 and t_2. Hence $N(t_1t_2)/(t_2 - t_1)$ will tend to the volume μ of T with $t_2 - t_1 \to \infty$. In computing that volume the multiplicity and orientation in which T covers the various parts of the torus are to be taken into account; the "characteristic function" of the trunk, to which the law of equidistribution (6) is here applied, is the covering index capable of the values $0, \pm 1, \pm 2, \cdots$. μ is perhaps more fittingly described as the flux of the constant velocity field λ through the slit, and is given by the integral over the slit of

$$\begin{vmatrix} \lambda_1, & \cdots, & \lambda_n \\ d_1\vartheta_1, & \cdots, & d_1\vartheta_n \\ \cdot & \cdots & \cdot \\ d_{n-1}\vartheta_1, & \cdots, & d_{n-1}\vartheta_n \end{vmatrix}. \tag{15}$$

$d_1, \cdots, d_{n-1}$ denote $n-1$ tangential line elements at a point (ϑ) of the slit in such orientation that (15) is positive or negative according as the current λ crosses the slit at (ϑ) in the positive or negative sense. We thus obtain

$$(7) \qquad \mu = W_1\lambda_1 + \cdots + W_n\lambda_n$$

where W_n is the area

$$W_n = \int \cdots \int d\vartheta_1 \cdots d\vartheta_{n-1}$$

of the vertical projection of the slit upon the $(n-1)$-dimensional $(\vartheta_1, \cdots, \vartheta_{n-1})$-torus with the orientation of its covering taken into account.

When we wish to ascertain whether (and with what index) a given point $(\vartheta_1 \cdots \vartheta_{n-1})$ is covered by the projection, we draw the $(n-1)$-linkage $A_0A_1 \cdots A_{n-1}$ with the links (3) starting at $A_0 = O$. Denoting A_{n-1} by b we have the figure as described before: the circle of radius $a_n = a$ around b and the origin O. We now add the negative real axis I issuing from O, and we have to watch whether the point $z(\vartheta)$ crosses I as ϑ varies from 0 to 1. $z(\vartheta)$ will cross once and in the positive sense if $|b| < a$; it will not cross at all, or once cross and then recross if $|b| > a$. Thus the covering index $= 1$ or 0 according to the distinction (12), and our previous formula (13) is confirmed.

When looked at in this way the topological situation is as plain as it could be. The flux of a constant velocity field through an $(n-1)$-dimensional surface bounded by an $(n-2)$-dimensional cycle does not depend on what two-sided surface one spans into the cycle; it may be described as the flux encompassed by the cycle and expressed by an integral extending over the cycle. If we had proceeded in this way we would have got entangled in a cobweb of topological difficulties. The singular manifold is of a complex topological structure which, moreover, varies with the values of the parameters a_k; the topologically different cases seem to be separated by those "exceptional" closed n-linkages whose a_k fall apart into two groups of equal sum. The unravelling of this maze is avoided by using the slit rather than the singular manifold, and by defining the slit precisely as we did.

By their very definition as probabilities the W_k are positive. But since we no longer make use of the Wintner integral (10) an independent proof of (14) becomes desirable. From the definition of $z(\vartheta_1 \cdots \vartheta_n)$ it is obvious that

$$\phi(\vartheta_1 + \vartheta, \cdots, \vartheta_n + \vartheta) = \vartheta + \phi(\vartheta_1, \cdots, \vartheta_n) \qquad (\vartheta \text{ arbitrary}),$$

and hence

$$\mu = \Sigma W_k\lambda_k$$

must increase by λ when all frequencies λ_k are raised by the same λ, or $\Sigma W_k = 1$. In arguing thus, one is a little bit encumbered by the assumption that the λ_k are linearly independent. But by choosing $\vartheta = -\vartheta_n$,

$$\phi(\vartheta_1, \cdots, \vartheta_n) = \phi(\vartheta_1 - \vartheta_n, \cdots, \vartheta_{n-1} - \vartheta_n, 0) + \vartheta_n;$$

one carries over the whole discussion from the n-dimensional to an $(n-1)$-dimensional phase space and finds the linear independence of $\lambda_1 - \lambda_n, \cdots, \lambda_{n-1} - \lambda_n$ to be a sufficient condition for our result. But this amounts to the absence of any relation (5) with integral coefficients m_k *of sum* 0, as stated in our theorem, and this hypothesis is not affected by passing from the frequencies λ_k to $\lambda_k + \lambda$.

So far we have assumed that the planet $z(t)$ does not pass through the origin. Should this happen at some moment t_0 the continuation of $\phi(t)$ beyond t_0 may become (and will in general become) ambiguous, with $\phi(t)$ splitting into two equally admissible branches which differ by 1. But our procedure shows that such an occurrence with its attendant accumulating ambiguity is rare enough to be of no influence upon the asymptotic law.

Combining the result (13) with (4) one gets for $n \geqq 3$ an explicit expression in terms of Bessel functions:

$$W_n = a_n \int_0^\infty J_1(a_n\rho) \prod_{k=1}^{n-1} J_0(a_k\rho)\, d\rho. \tag{16}$$

I conclude with a few remarks on the history of our question. When Lagrange hit upon it in his approximate theory of secular perturbations, he at once solved the "trivial" case where one of the sides, a_n say, is larger than the sum of all others,

$$a_n > a_1 + \cdots + a_{n-1}.$$

One then has $\mu = \lambda_n$ with the much sharper estimate

$$\phi(t) = \lambda_n t + O(1).$$

(This estimate was implied in the term "mean motion," as Lagrange used it. Actually all planets are in the Lagrangean case except Venus and Earth.) More than a century elapsed before P. Bohl succeeded in establishing mean motion for $n = 3$ in the non-Lagrangean case where a_1, a_2, a_3 form a triangle Δ. He found

$$\mu = W_1\lambda_1 + W_2\lambda_2 + W_3\lambda_3,$$

πW_k being the angles of Δ. This result is easily obtained if after the aforementioned elementary reduction of the dimensionality of the phase space to $n-1$, one expresses the flux in terms of the singular manifold which now is a 0-cycle. When the problem and Bohl's paper were pointed out to me by Felix Bernstein in 1913, it started me on my investigations on Diophantine approximations (W_1). Although the essential step was accomplished by the equidistribution law, the messy topology of the singular manifold at that time prevented me from settling the problem by an explicit and universal formula. As an illustration I carried out the case $n = 4$; the topology of closed 4-linkages has been thoroughly investigated in kinematics because such linkages are a frequent element in the construction of machines. Even there I limited myself to one of the topological subcases, see W_1, Satz 7, and blundered in the final formulation (although the proof is correct): since my "integral invariants" of the rickety quadrilateral have the sum 0 instead of 1, the term λ_1 should be added in the final formula. But I will not mar the simplicity of our present argument by these old vagaries. Awareness of those complications will, however, help the reader to a fuller appreciation of the result (8), (16) holding for all cases alike.

This formula, which served me as a lodestar, was guessed for all n and verified for $n = 3, 4$ by Professor Wintner. His report on the paper HKW in my Princeton seminar on current literature stimulated me to resume the ancient problem. The various equations known about Bessel integrals of the type (16) for lower n are most naturally derived from the interpretation put upon (16) by the problem of mean motion.

The Institute for Advanced Study,
Princeton, N. J.

BIBLIOGRAPHY.

[1] See e. g. Charlier, *Die Mechanik des Himmels* I, Leipzig, 1902, pp. 333-436, and for a brief account the paper W_2 by the author, *Enseignement Mathématique*, vol. 16 (1914), p. 455.
[2] A. Wintner, *American Journal of Mathematics*, vol. 55 (1933), p. 309.
[3] Cf. Watson's *Treatise on Bessel Functions*, pp. 411, 414, and 420.
[4] See the paper W_1 by the author, *Mathematische Annalen*, vol. 77 (1916), p. 313.
[5] *American Journal of Mathematics*, vol. 59 (1937), p. 261, cited as HKW.
[6] P. Bohl, *Journ. reine u. angew. Mathem.*, vol. 135 (1909), p. 189.

THE HOMOGENEOUS CHAOS.

By Norbert Wiener.

1. Introduction. Physical need for theory.
2. Definition. Types of chaos.
3. Classical ergodic theorem. Lebesgue form.
4. Dominated ergodic theorem. Multidimensional ergodic theorem.
5. Metric transitivity. Space and phase averages in a chaos.
6. Pure one-dimensional chaos.
7. Pure multidimensional chaos.
8. Phase averages in a pure chaos.
9. Forms of chaos derivable from a pure chaos.
10. Chaos theory and spectra.
11. The discrete chaos.
12. The weak approximation theorem for the polynomial chaos.
13. The physical problem. The transformations of a chaos.

1. Introduction. Physical need for theory. Statistical mechanics may be defined as the application of the concepts of Lebesgue integration to mechanics. Historically, this is perhaps putting the cart before the horse. Statistical mechanics developed through the entire latter half of the nineteenth century before the Lebesgue integral was discovered. Nevertheless, it developed without an adequate armory of concepts and mathematical technique, which is only now in the process of development at the hands of the modern school of students of integral theory.

In the more primitive forms of statistical mechanics, the integration or summation was taken over the manifold particles of a single homogeneous dynamical system, as in the case of the perfect gas. In its more mature form, due to Gibbs, the integration is performed over a parameter of distribution, numerical or not, serving to label the constituent systems of a dynamical ensemble, evolving under identical laws of force, but differing in their initial conditions. Nevertheless, the study of the mode in which this parameter of distribution enters into the individual systems of the ensemble does not seem to have received much explicit study. The parameter of distribution is essentially a parameter of integration only. As such, questions of dimensionality are indifferent to it, and it may be replaced by a numerical variable α with the range $(0, 1)$. Any transformation leaving invariant the probability properties of the ensemble as a whole is then represented by a measure-preserving transformation of the interval $(0, 1)$ into itself.

Among the simplest and most important ensembles of physics are those which have a spatially homogeneous character. Among these are the homogeneous gas, the homogeneous liquid, the homogeneous state of turbulence. In these, while the individual systems may not be invariant under a change of origin, or, in other words, under the translation of space by a vector, the ensemble as a whole is invariant, and the individual systems are merely permuted without change of probability. From what we have said, the translations of space thus generate an Abelian group of equi-measure transformations of the parameter of distribution.

One-dimensional groups of equi-measure transformations have become well known to the mathematicians during the past decade, as they lie at the root of Birkhoff's famous ergodic theorem.[1] This theorem asserts that if we have given a set S of finite measure, an integrable function $f(P)$ on S, and a one-parameter Abelian group T^λ of equi-measure transformations of S into itself, such that

$$(1) \qquad T^\lambda \cdot T^\mu = T^{\lambda+\mu} \qquad (-\infty < \lambda < \infty, \; -\infty < \mu < \infty),$$

then for all points P on S except those of a set of zero measure, and provided certain conditions of measurability are satisfied,

$$(2) \qquad \lim_{A\to\infty} \frac{1}{A} \int_0^A f(T^\lambda P)\, d\lambda$$

will exist. Under certain more stringent conditions, known as metric transitivity, we shall have

$$(3) \qquad \lim_{A\to\infty} \frac{1}{A} \int_0^A f(T^\lambda P)\, d\lambda = \int_S f(P)\, dV_P$$

almost everywhere. The ergodic theorem thus translates averages over an infinite range, taken with respect to λ, into averages over the set S of finite measure. Even without metric transitivity, the ergodic theorem translates the distribution theory of λ averages into the theory of S averages.

In the most familiar applications of the ergodic theorem, S is taken to be a spatial set, and the parameter λ is identified with the time. The theorem thus becomes a way of translating time averages into space averages, in a manner which was postulated by Gibbs without rigorous justification, and which forms the entire basis of his methods. Strictly speaking, the space averages are generally in phase-space rather than in the ordinary geometrical space of three dimensions. There is no reason, however, why the parameter λ should be confined to one taking on values of the time for its arguments, nor even why it should be a one-dimensional variable. We are thus driven to formulate and prove a multidimensional analogue of the classical Birkhoff theorem.

---

[1] Cf. Eberhard Hopf, "Ergodentheorie," *Ergebnisse der Mathematik und ihrer Grenzgebiete*, vol. 5. See particularly § 14, where further references to the literature are given.

In the ordinary Birkhoff theorem, the transformations T^λ are taken to be one-one point transformations. Now, the ergodic theorem belongs fundamentally to the abstract theory of the Lebesgue integral, and in this theory, individual points play no rôle. In the study of chaos, individual values of the parameter of integration are equally unnatural as an object of study, and it becomes desirable to recast the ergodic theorem into a true Lebesgue form. This we do in paragraph 2.

Of all the forms of chaos occurring in physics, there is only one class which has been studied with anything approaching completeness. This is the class of types of chaos connected with the theory of the Brownian motion. In this one-dimensional theory, there is a simple and powerful algorithm of phase averages, which the ergodic theorem readily converts into a theory of averages over the transformation group. This theory is easily generalized to spaces of a higher dimensionality, without any very fundamental alterations. We shall show that there is a certain sense in which these types of chaos are central in the theory, and allow us to approximate to all types.

Physical theories of chaos, such as that of turbulence, or of the statistical theory of a gas or a liquid, may or may not be theories of equilibrium. In the general case, the statistical state of a chaotic system, subject to the laws of dynamics, will be a function of the time. The laws of dynamics produce a continuous transformation group, in which the chaos remains a chaos, but changes its character. This is at least the case in those systems which can continue to exist indefinitely in time without some catastrophe which essentially changes their dynamic character. The study, for example, of the development of a state of turbulence, depends on an existence theory which avoids the possibility of such a catastrophe. We shall close this paper by certain very general considerations concerning the demands of an existence theory of this sort.

2. Definition. Types of Chaos. A *continuous homogeneous chaos* in n dimensions is a scalar or vector valued measurable function $\rho(x_1, \cdots, x_n; \alpha)$ of $x_1, \cdots, x_n; \alpha$, in which $x_1, \cdots, x_n$ assume all real values, while α ranges over $(0, 1)$; and in which the set of values of α for which

$$\rho(x_1+y_1,\cdots,x_n+y_n;\alpha) \text{ belongs to } S, \tag{4}$$

if it has a measure for any set of values of $y_1,\cdots,y_n$, has the same measure for any other set of values. In this paper, we shall confine our attention to scalar chaoses. A continuous homogeneous chaos is said to be *metrically transitive,* if whenever the sets of values of α for which $\rho(x_1,\cdots,x_n;\alpha)$ belongs to S and to S_1, respectively, have measures M and M_1, the set of values of α for which simultaneously

$$\rho(x_1,\cdots,x_n;\alpha) \text{ belongs to } S$$

and

$$\rho(x_1+y_1,\cdots,x_n+y_n;\alpha) \text{ belongs to } S_1$$

has a measure which tends to M_1 as $y_1^2+\cdots+y_n^2\to\infty$.

If ρ is integrable, it determines the additive set-function

$$\mathfrak{F}(\Sigma;\alpha)=\int\cdots_{\Sigma}\int\rho(x_1,\cdots,x_n;\alpha)\,dx_1\cdots dx_n. \tag{5}$$

On the other hand, not every additive set-function may be so defined. This suggests a more general definition of a *homogeneous chaos,* in which the chaos is defined to be a function $\mathfrak{F}(\Sigma;\alpha)$, where α ranges over $(0,1)$ and Σ belongs to some additively closed set Ξ of measurable sets of points in n-space. We suppose that if Σ and Σ_1 do not overlap,

$$\mathfrak{F}(\Sigma+\Sigma_1;\alpha)=\mathfrak{F}(\Sigma;\alpha)+\mathfrak{F}(\Sigma_1;\alpha). \tag{6}$$

We now define the new point-set $\Sigma(y_1,\cdots,y_n)$ by the assertion

$$\Sigma(y_1,\cdots,y_n) \text{ contains } x_1+y_1,\cdots,x_n+y_n \text{ when and only when } \Sigma \text{ contains } x_1,\cdots,x_n. \tag{7}$$

This leads to the definition of the additive set-function $\mathfrak{F}_{y_1,\ldots,y_n}(\Sigma;\alpha)$ by

$$\mathfrak{F}_{y_1,\ldots,y_n}(\Sigma;\alpha)=\mathfrak{F}(\Sigma(y_1,\cdots,y_n);\alpha). \tag{8}$$

If, then, for all classes S of real numbers,

$$\text{Measure of set of } \alpha\text{'s for which } \mathfrak{F}_{y_1,\ldots,y_n}(\Sigma;\alpha) \text{ belongs to class } S \tag{9}$$

is independent of $y_1,\cdots,y_n$, in the sense that if it exists for one set of these numbers, it exists for all sets, and has the same value, and if it is measurable in $y_1,\cdots,y_n$, we shall call $\mathfrak{F}$ a homogeneous chaos. The notion of metrical transitivity is generalized in the obvious way, replacing $\rho(x_1,\cdots,x_n;\alpha)$ by $\mathfrak{F}(\Sigma;\alpha)$, and $\rho(x_1+y_1,\cdots,x_n+y_n;\alpha)$ by $\mathfrak{F}_{y_1,\ldots,y_n}(\Sigma;\alpha)$.

The theorem which we wish to prove is the following:

THEOREM I. *Let $\mathfrak{F}(\Sigma;\alpha)$ be a homogeneous chaos. Let the functional*

$$\Phi\{\mathfrak{F}(\Sigma;\alpha)\} = g(\alpha) \tag{10}$$

be a measurable function of α, such that $\int_0^1 |\, g(\alpha) \log^+ |\, g(\alpha)|\,|\, d\alpha$ is finite. Then for almost all values of α,

$$\lim_{r\to\infty} \frac{1}{V(r)} \int \cdots_R \int \Phi\{\mathfrak{F}_{y_1,\ldots,y_n}(\Sigma;\alpha)\} dy_1 \cdots dy_n \tag{11}$$

exists, where R is the interior of the sphere

$$y_1^2 + y_2^2 + \cdots + y_n^2 = r^2 \tag{12}$$

and $V(r)$ is its volume. If in addition, $\mathfrak{F}(\Sigma;\alpha)$ is metrically transitive,

$$\lim_{r\to\infty} \frac{1}{V(r)} \int \cdots_R \int \Phi\{\mathfrak{F}_{y_1,\ldots,y_n}(\Sigma;\alpha)\} dy_1 \cdots dy_n = \int_0^1 \Phi\{\mathfrak{F}(\Sigma;\beta)\} d\beta \tag{13}$$

for almost all values of α.

3. Classical ergodic theorem. Lebesgue form. Theorem I is manifestly a theorem of the ergodic type. Let it be noticed, however, that we nowhere assume that the transformation of α given by $\beta = T\alpha$ when

$$\mathfrak{F}_{y_1,\ldots,y_n}(\Sigma;\beta) = \mathfrak{F}(\Sigma;\alpha) \tag{14}$$

is one-one. This should not be surprising, as the ergodic theorem is fundamentally one concerning the Lebesgue integral, and in the theory of the Lebesgue integral, individual points play no rôle.

Nevertheless, in the usual formulation of the ergodic theorem, the expression $f(T^\lambda P)$ enters in an essential way. Can we give this a meaning without introducing the individual transform of an individual point?

The clue to this lies in the definition of the Lebesgue integral itself. If $f(P)$ is to be integrated over a region S, we divide S into the regions $S_{a,b}(f)$, defined by the condition that over such a region,

$$a < f(P) \leq b. \tag{15}$$

We now write

$$\int_S f(P)\, dV_P = \lim_{\epsilon\to 0} \sum_{-\infty}^{\infty} n\epsilon\, m(S_{(n-1)\epsilon,\, n\epsilon}(f)). \tag{16}$$

The condition that $f(P)$ be integrable thus implies the condition that it be measurable, or that all the sets $S_{a,b}(f)$ be measurable.

Now, if T is a measure-preserving transformation on S, the sets $TS_{a,b}(f)$

will all be measurable, and will have, respectively, the same measures as the sets $S_{a,b}(f)$. We shall define the function $f(TP) = g(P)$ by the conditions

$$TS_{a,b}(f) = S_{a,b}(g). \tag{17}$$

If T conserves relations of inclusion of sets, up to sets of zero measure, this function will clearly be defined up to a set of values of P of zero measure, and we shall have

$$\int_S f(TP)\,dV_P = \int_S f(P)\,dV_P. \tag{18}$$

We may thus formulate the original or discrete case of the Birkhoff ergodic theorem, as follows: *Let S be a set of points of finite measure. Let T be a transformation of all measurable sub-sets of S into measurable sub-sets of S, which conserves measure, and the relation between two sets, that one contains the other except at most for a set of zero measure. Then except for a set of points P of zero measure,*

$$\lim_{N\to\infty} \frac{1}{N+1} \cdot \sum_0^N f(T^n P) \tag{19}$$

will exist.

The continuous analogue of this theorem needs to be formulated in a somewhat more restricted form, owing to the need of providing for the integrability of the functions concerned. It reads: *Let S be a set of points of finite measure. Let T^λ be a group of transformations fulfilling the conditions we have laid down for T in the discrete case just mentioned. Let $T^\lambda P$ be measurable in the product space of λ and of P. Then, except for a set of points P of zero measure,*

$$\lim_{A\to\infty} \frac{1}{A} \int_0^A f(T^\lambda P)\,d\lambda \tag{20}$$

will exist.

In the proofs of Birkhoff's ergodic theorem, as given by Khintchine and Hopf, no actual use is made of the fact that the transformation T is one-one, and the proofs extend to our theorem as stated here, without any change. The restriction of measurability, or something to take its place, is really necessary for the correct formulation of Khintchine's statement of the ergodic theorem, as Rademacher, von Neumann, and others have already pointed out.[2]

With the aid of the proper Lebesgue formulation of the ergodic theorem, the one-dimensional case of Theorem I follows at once. Actually more follows, as it is only necessary that g belong to L, instead of to the logarithmic class with which we replace it. To prove Theorem I in its full generality, we must establish a multidimensional ergodic theorem.

---

[2] Cf. J. v. Neumann, *Annals of Mathematics*, 2, vol. 33, p. 589, note 11.

4. Dominated ergodic theorem. Multidimensional ergodic theorem. As a lemma to the multidimensional ergodic theorem, we first wish to establish the fact that if the function $f(P)$ in the ergodic theorem satisfies the condition

$$\int_S |f(P)| \log^+ |f(P)|\, dV_P < \infty, \tag{21}$$

then the expressions (19) and (20) not only exist, but the limits in question will be approached under the domination of a summable function of P. We shall prove this in the discrete case, for the sake of simplicity, but the result goes over without difficulty to the continuous case.

Let T be an equimeasure transformation of the set S of finite measure into itself, in the generalized sense of the last paragraph, and let W be a measurable sub-set of S, with the characteristic function $W(P)$. Let U be the set of all points P for which some $T^{-j}P$ belongs to W. Let $i(P)$, when P belongs to W, be defined as the smallest positive number n such that $T^{-n}P$ belongs to W, and let us call it the *index* of P. Every point of W, except for a set of measure 0, will have a finite index, since if we write W_∞ for the set of points without a finite index, no two sets $T^m W_\infty$ and $T^n W_\infty$ can overlap, while they all have the same measure and their sum has a finite measure. Thus except for a set of zero measure, we may divide W into the sets W_p, each consisting of the points of W of index p. It is easy to show that if $1 \leq p < \infty$, $1 \leq p' < \infty$, $0 \leq j < p$, $0 \leq j' < p'$, the sets $T^{-j}W_p$ and $T^{-j'}W_{p'}$ can not overlap over a set of positive measure unless $j = j'$, $p = p'$. Similarly, the sets $T^{-p}W_p$ and $T^{-p'}W_{p'}$ can not overlap over sets of positive measure, unless $p = p'$, and represent a dissection of W, except for a set of zero measure. Let us put W_{pq} for the logical product of W_p and $T^p W_q$; W_{pqr} for the logical product of W_{pq} and $T^{p+q}W_r$; and so on. Then if k is fixed, the sets $T^{-j}W_{p_1p_2\ldots p_k}$ cover U (except for sets of zero measure) once as j goes from 0 to p_1, once as it goes from p_1 to $p_1 + p_2$, and so on; making just k times between 0, inclusive, and $p_1 + p_2 + \cdots + p_k$, exclusive. Thus if S_K is the set of all the points in all the $W_{p_1p_2\ldots p_k}$ for which $p_1 + p_2 + \cdots + p_k = K$, the total measure of all the S_K's for $2^{N+1} \geq K \geq 2^N$ can not exceed $km(S)/2^N$.

Now let P lie in $T^{-j}W_{p_1p_2\ldots p_k}$, where $0 \leq j < p_1 + p_2 + \cdots + p_k$. Let us consider the sequence of numbers a_j, where if $p_1 + \cdots + p_l \leq j < p_1 + \cdots + p_{l+1}$, a_j is the greatest of the numbers

$$\frac{1}{j - p_1 - \cdots - p_l + 1}, \quad \frac{2}{j - p_1 - \cdots - p_{l-1} + 1}, \quad \frac{3}{j - p_1 - \cdots - p_{l-2} + 1}, \quad \cdots, \quad \frac{l+1}{j+1}.$$

Then a_j will be the largest of the numbers

$$W(P),\ \frac{W(P)+W(TP)}{2},\cdots,\frac{W(P)+W(TP)+\cdots+W(T^jP)}{j+1}.$$

The sum $\sum_0^k a_j$, for a fixed K, will have its maximum value when $p_1 = p_2 = \cdots = p_{k-1} = 1$; $p_k = K+1-k$, when it will be $k + \sum_{k+1}^{k} k/j \leq k(1+\log K/k)$. In this case the sequence of the a_j's will be

$$\underbrace{1, 1, \cdots, 1,}_{k \text{ times}}\ \frac{k}{k+1}, \frac{k}{k+2}, \cdots, \frac{k}{K}.$$

This remark will be an easy consequence of the following fact: let us consider the sequence

$$\cdots,\frac{\lambda}{n-1},\frac{\lambda}{n},1,\frac{1}{2},\cdots,\frac{1}{\left[\frac{n}{\lambda}\right]},\frac{\lambda+1}{\left[\frac{n}{\lambda}\right]+n+1},\cdots \tag{22}$$

and the modified sequence

$$\cdots,\frac{\lambda}{n-1},1,\frac{1}{2},\cdots, \\ \frac{1}{\left[\frac{n-1}{\lambda}\right]},\frac{\lambda+1}{\left[\frac{n-1}{\lambda}\right]+n},\frac{\lambda+1}{\left[\frac{n-1}{\lambda}\right]+n+1},\cdots \tag{23}$$

where of course

$$\frac{\lambda+1}{n+\left[\frac{n}{\lambda}\right]}\leq\frac{1}{\left[\frac{n}{\lambda}\right]};\quad \frac{\lambda+1}{n+\left[\frac{n}{\lambda}\right]+1}>\frac{1}{\left[\frac{n}{\lambda}\right]+1};$$

$$\frac{\lambda+1}{n-1+\left[\frac{n-1}{\lambda}\right]}\leq\frac{1}{\left[\frac{n-1}{\lambda}\right]};\quad \frac{\lambda+1}{n+\left[\frac{n-1}{\lambda}\right]}>\frac{1}{\left[\frac{n-1}{\lambda}\right]+1}.$$

Apart from the arrangement, the terms of (22) will be the same as the terms of (23), except that in (23), $\dfrac{\lambda+1}{\left[\frac{n-1}{\lambda}\right]+n}$ replaces λ/n. Now,

$$\frac{\lambda+1}{\left[\frac{n-1}{\lambda}\right]+n}>\frac{\lambda}{n},$$

so that the sum of the terms in (23) is greater than that in (22).

Since the transforms of a given set have the same measure as the set, and the sets $T^{-j}S_K$ cover U exactly k times, we have

$$\begin{aligned}
\frac{1}{k}\int_S \Big\{ & W(P) + \max\left(W(P), \frac{W(P)+W(TP)}{2}\right) \\
& + \max\left(W(P), \frac{W(P)+W(TP)}{2}, \frac{W(P)+W(TP)+W(T\,P)}{3}\right) \\
& + \cdots + \max\left(W(P), \frac{W(P)+W(TP)}{2}, \cdots,\right. \\
& \qquad\qquad \left.\frac{W(P)+\cdots+W(T^{k-1}P)}{k}\right)\Big\}\, dV_P \\
& \leq \sum_{K=0}^{\infty} m(S_K)\left(1+\log\frac{K}{k}\right) \\
& \leq m(W)\left(1+\log\frac{2^N}{k}\right) + \sum_{j=1}^{\infty}\frac{km(S)}{2^{N+j}}\left(1+\log\frac{2^{N+j}}{k}\right) \\
& \leq \left(1+\log\frac{2^N}{k}\right)\left\{ m(W) + \text{const.}\,\frac{km(S)}{2^N}\right\},
\end{aligned}$$

the constant being absolute. If we now put

$$N = \left[\log\frac{km(S)}{m(W)}\Big/\log 2\right],$$

this last expression is dominated by

$$\text{const.}\, m(W)\left(1+\log\frac{m(S)}{m(W)}\right).$$

Since the constant is independent of k, we see that the Cesàro average of

$$\max\left(W(P), \frac{W(P)+W(TP)}{2}, \cdots, \frac{W(P)+\cdots+W(T^{k-1}P)}{k}\right).$$

is dominated by the same term. Thus

$$\begin{aligned}
\int_S \max\left(W(P), \cdots, \frac{W(P)+\cdots+W(T^{k-1}P)}{k}\right) dV_P \\
\leq \text{const.}\, m(W)\left(1+\log\frac{m(S)}{m(W)}\right),
\end{aligned}$$

and it follows by monotone convergence that there exists a function $W^*(P)$ such that

$$\int_S W^*(P)\,dV_P \leq \text{const.}\, m(W)\left(1+\log\frac{m(S)}{m(W)}\right), \tag{24}$$

and for all positive m,

$$\frac{W(P)+\cdots+W(T^mP)}{m+1} \leq W^*(P). \tag{25}$$

Now let $f(P)$ be a function such that

$$\int_S |f(P)| \log^+ |f(P)| \, dV_P < \infty. \tag{21}$$

Let $W^{(N)}$ be the set of points such that

$$2^N \leq |f(P)| \leq 2^{N+1}, \tag{26}$$

and let $f^{(N)}(P)$ be the function equal to $f(P)$ over this set of points, and 0 elsewhere. Then there exists a function $f^{(N)*}(P)$, such that

$$\frac{f^{(N)}(P) + \cdots + f^{(N)}(T^m P)}{m+1} \leq f^{(N)*}(P) \qquad (m = 0, 1, 2, \cdots), \tag{27}$$

and

$$\int_S f^{(N)*}(P) \, dV_P \leq \text{const.}\ m(W^{(N)}) \left(1 + \log^+ \frac{m(S)}{m(W^{(N)})}\right) 2^N. \tag{28}$$

Hence if

$$\sum_{-\infty}^{\infty} m(W^{(N)}) \left(1 + \log^+ \frac{m(S)}{m(W^{(N)})}\right) 2^N < \infty \tag{29}$$

and

$$f^*(P) = \sum_{-\infty}^{\infty} f^{(N)*}(P), \tag{30}$$

then

$$\frac{f(P) + \cdots + f(T^m P)}{m+1} \leq f^*(P) \qquad (m = 0, 1, 2, \cdots), \tag{31}$$

and

$$\int_S f^*(P) \, dV_P \leq \text{const.} \sum m(W^{(N)}) \left(1 + \log^+ \frac{m(S)}{m(W^{(N)})}\right) 2^N. \tag{32}$$

However,

$$\begin{aligned} &\sum_{-\infty}^{\infty} m(W^{(N)}) \left(1 + \log^+ \frac{m(W^{(N)})}{m(S)}\right) 2^N \\ &\qquad \leq \sum_{-\infty}^{\infty} \int_{W^{(N)}} |f(P)| \, dV_P \left(1 + \log^+ \frac{2^{N+1} m(S)}{\int_S |f(P)| \, dV_P}\right) \\ &\qquad \leq \int_S |f(P) \left(1 + \log^+ \frac{2\, |f(P)|\, m(S)}{\int_S |f(Q) \, dV_Q}\right) dV_P \\ &\qquad \leq \int_S |f(P)| \, dV_P \left(1 + \log^+ \frac{2m(S)}{\int_S |f(P)| \, dV_P}\right) \\ &\qquad\qquad + \int_S |f(P)| \log^+ |f(P)| \, dV_P. \end{aligned} \tag{33}$$

Thus $\int_S f^*(P)\,dV_P$ has an upper bound which is less than a function of $\int_S |f(P)|\,dV_P$ and $\int_S |f(P)|\log^+|f(P)|\,dV_P$, tending to 0 as they both tend to 0. This establishes our theorem of the existence of a uniform dominant.

There is a sense in which (21) *is a best possible condition.* That is, if

$$\psi(x) = o(\log^+ x), \tag{34}$$

the condition

$$\int_S (\psi(|f(P)|)\,|f(P)|\,dV_P < \infty \tag{35}$$

is not sufficient for the existence of a uniform dominant. For let S be a set of measure 1, subdivided into mutually exclusive sets S_n, of measures respectively 2^{-n}. Let S_n be divided into mutually exclusive sets $S_{n,1}, \cdots S_{n,\nu_n}$, all of equal measures. Let T transform $S_{n,k}$ into $S_{n,(k+1)}$ $(k < \nu^n)$, and S_{n,ν_n} into $S_{n,1}$. Let $f(P)$ be defined by

$$f(P) = a_n > 0 \text{ on } S_{n_1}, \quad (n = 1, 2, \cdots); \quad f(P) = 0 \text{ elsewhere.} \tag{36}$$

Then the smallest possible uniform dominant of

$$\frac{1}{N+1}\sum_0^N f(T^n P) \tag{37}$$

is

$$f^*(P) = \frac{a_n}{k} \text{ on } S_{n,k}, \tag{38}$$

and we have

$$\int_S f^*(P)\,dV_P = \sum \frac{a_n \Omega(\log \nu_n) 2^{-n}}{\nu_n}. \tag{39}$$

Thus if

$$\nu_n = 2^{2n} a_n, \quad a_n = \Omega(n!), \tag{40}$$

the function $f^*(P)$ will belong to L if and only if

$$\infty > \sum \frac{a_n \Omega(\log a_n) 2^{-n}}{\nu_n} > \text{const.} \int_S f(P) \log^+ f(P)\,dV_P. \tag{41}$$

While we have proved the dominated ergodic theorem merely as a lemma for the multidimensional ergodic theorem, the theorem, and more particularly the method by which we have proved it, have very considerable independent interest. We may use these methods to deduce von Neumann's mean ergodic theorem from the Birkhoff theorem; or vice versa, we may deduce the Birkhoff theorem, at least in the case of a function satisfying (21), from the von

Neumann theorem. These facts however are not relevant to the frame of the present paper, and will be published elsewhere.

We shall now proceed to the proof of the multidimensional ergodic theorem, which we shall establish in the two-dimensional case, although the method is independent of the number of dimensions. Let $T_1^{\lambda} T_2^{\mu}$ be a two-dimensional Abelian group of transformations of the set S (of measure 1) into itself, in the sense in which we have used this term in paragraph 3. Let $T_1^{\lambda} T_2^{\mu} P$ be measurable in λ, μ, and P. We now introduce a new variable x, ranging over $(0, 1)$, and form the product space Σ of P and x. We introduce the one-parameter group of transformations of this space, T^{ρ}, by putting

$$T^{\rho}(P, x) = (T_1^{\rho \cos 2\pi x} T_2^{\rho \sin 2\pi x} P, x) \tag{42}$$

The expressions $T^{\rho}(P, x)$ will be measurable in ρ and (P, x), and the transformations T^{ρ} will all preserve measure on Σ. Thus by the ergodic theorem, for almost all points (P, x) of Σ, if $f(P) = f(P, x)$ belongs to L, the limit

$$\lim_{A \to \infty} \frac{1}{A} \int_0^A f(T^{\rho}(P, x))\, d\rho = \lim_{A \to \infty} \frac{1}{A} \int_0^A f(T_1^{\rho \cos 2\pi x} T_2^{\rho \sin 2\pi x} P)\, d\rho \tag{43}$$

will exist. If condition (26) is satisfied, it will follow by dominated convergence that the limit

$$\lim_{A \to \infty} \frac{1}{A} \int_0^A d\rho \int_0^1 f(T_1^{\rho \cos 2\pi x} T_2^{\rho \sin 2\pi x} P)\, dx \tag{44}$$

will exist for almost all points (P, x), and hence for almost all points P.

For the moment, let us assume that $f(P)$ is non-negative. Then there is a Tauberian theorem, due to the author,[3] which establishes that the expression (44) is equivalent to

$$\lim_{A \to \infty} \frac{1}{\pi A^2} \int_0^A \rho\, d\rho \int_0^{2\pi} f(T_1^{\rho \cos \theta} T_2^{\rho \sin \theta} P)\, d\theta. \tag{45}$$

The only point of importance which we must establish in order to justify this Tauberian theorem is that

$$\int_0^1 \rho^{1+iu}\, d\rho = \frac{1}{2 + iu} \neq 0 \tag{46}$$

for real values of u. Since every function $f(P)$ satisfying (21) is the difference of two non-negative functions satisfying this condition, (45) is established in the general case.

It will be observed that we have established our multidimensional ergodic theorem on the basis of assumption (21), and not on that of the weaker assumption that $f(P)$ belongs to L. What the actual state of affairs may be, we do not know. At any rate, all attempts to arrive at a direct analogue of the Khintchine proof for one dimension have broken down. The one-dimensional proof makes essential use of the fact that the difference of two intervals is always an interval, while the difference between two spheres is not always a sphere.

---

[3] N. Wiener, "Tauberian theorems," *Annals of Mathematics*, 2, vol. 33, p. 28.

The precise statement of the multidimensional ergodic theorem is the following: *Let S be a set of points of measure* 1, *and let $T_1^{\lambda_1}T_2^{\lambda_2}\cdots T_n^{\lambda_n}$ be an Abelian group of equimeasure transformations of S into itself, in the sense of paragraph* 3. *Let $T_1^{\lambda_1}\cdots T_n^{\lambda_n}P$ be measurable in $\lambda_1,\cdots,\lambda_n$; P. Let R be the set of values of $\lambda_1,\cdots,\lambda_n$ for which*

$$\lambda_1^2+\lambda_2^2+\cdots+\lambda_n^2\leq r^2 \tag{47}$$

and let $V(r)$ be its volume. Let $f(P)$ satisfy the condition (21). *Then for almost all values of P,*

$$\lim_{r\to\infty}\frac{1}{V(r)}\int\underset{R}{\cdots}\int f(T_1^{\lambda_1}\cdots T_n^{\lambda_n}P)\,d\lambda_1\cdots d\lambda_n \tag{48}$$

exists.

That part of Theorem I which does not concern metric transitivity is an immediate corollary.

5. Metric transitivity. Space and Phase Averages in a Chaos. If the function $f(P)$ is positive, clearly

$$\begin{aligned}
\int\cdots\int_{\lambda_1^2+\cdots+\lambda_n^2\leq r^2-\mu_1^2-\cdots-\mu_n^2} & f(T_1^{\lambda}\cdots T_n^{\lambda}P)\,d\lambda_1\cdots d\lambda_n \\
&\leq\int\cdots\int_{\lambda_1^2+\cdots+\lambda_n^2\leq r^2} f(T_1^{\lambda_1}\cdots T_n^{\lambda_n}T_1^{\mu_1}\cdots T_n^{\mu_n}P)\,d\lambda_1\cdots d\lambda_n \\
&\leq\int\cdots\int_{\lambda_1^2+\cdots+\lambda_n^2\leq r^2+\mu_1^2+\cdots+\mu_n^2} f(T_1^{\lambda_1}\cdots T_n^{\lambda_n}P)\,d\lambda_1\cdots d\lambda_n.
\end{aligned} \tag{49}$$

Hence

$$\begin{aligned}
\lim_{r\to\infty}\frac{1}{V(r)}\int\underset{R}{\cdots}\int & f(T_1^{\lambda_1}\cdots T_n^{\lambda_n}T_1^{\mu_1}\cdots T_n^{\mu_n}P)\,d\lambda_1\cdots d\lambda_n \\
&=\lim_{r\to\infty}\frac{1}{V(r)}\int\underset{R}{\cdots}\int f(T_1^{\lambda_1}\cdots T_n^{\lambda_n}P)\,d\lambda_1\cdots d\lambda_n,
\end{aligned} \tag{50}$$

and expression (48) has the same value for P and all its transforms under the group $T_1^{\lambda_1}\cdots T_n^{\lambda_n}$. The condition of positivity is clearly superfluous. Thus in case expression (48) does not almost everywhere assume a single value, there will be two classes S_1 and S_2 of elements of S, each of positive measure, and each invariant under all the transformations $T_1^{\lambda_1}\cdots T_n^{\lambda_n}$.

A condition which will manifestly exclude such a contingency is that if S_1 and S_2 are two sub-sets of S of positive measure, and ϵ is a positive quantity, there always exists a transformation $T = T_1^{\lambda_1}\cdots T_n^{\lambda_n}$, such that

$$(51)\qquad \left| \frac{mS_1(TS_2)}{mS_2} - mS_1 \right| < \epsilon.$$

From this it will immediately follow that if a chaos is metrically transitive in the sense of paragraph 3, the group of transformations of the α space generated by translations of the chaos will have the property we have just stated, and under the assumptions of Theorem I,

$$(52)\qquad \lim_{r\to\infty} \frac{1}{V(r)} \int \cdots_{R} \int \Phi\{\mathfrak{F}_{y_1,\dots,y_n}(\Sigma;\alpha)\}\,dy_1\cdots dy_n$$

will exist and have the same value for almost all values of α.

If almost everywhere

$$(53)\qquad \lim_{r\to\infty} \frac{1}{V(r)} \int \cdots_{R} \int \Phi\{\mathfrak{F}_{y_1,\dots,y_n}(\Sigma;\alpha)\}\,dy_1\cdots dy_n = A,$$

and

$$(54)\qquad \frac{1}{V(r)} \int \cdots_{R} \int \Phi\{\mathfrak{F}_{y_1,\dots,y_n}(\Sigma;\alpha)\}\,dy_1\cdots dy_n < g(\alpha)$$

where $g(\alpha)$ belongs to L, then by dominated convergence,

$$(55)\qquad A = \lim_{r\to\infty} \frac{1}{V(r)} \int \cdots_{R} \int dy_1\cdots dy_n \int_0^1 d\alpha\,\Phi\{\mathfrak{F}_{y_1,\dots,y_n}(\Sigma;\alpha)\}$$
$$= \int_0^1 \Phi\{\mathfrak{F}(\Sigma;\alpha)\}\,d\alpha.$$

That is, the average of $\Phi\{\mathfrak{F}(\Sigma;\alpha)\}$, taken over the finite phase space of α, is almost everywhere the same as the average of $\Phi\{\mathfrak{F}_{y_1,\dots,y_n}(\Sigma;\alpha)\}$ taken over the infinite group space of points $y_1,\cdots,y_n$. This completes the establishment of Theorem I, and gives us a real basis for the study of the homogeneous chaos.[4]

6. Pure one-dimensional chaos. The simplest type of pure chaos is that which has already been treated by the author in connection with the Brownian motion. However, as we wish to generalize this theory to a multiplicity of dimensions, instead of referring to existing articles on the subject, we shall present it in a form which emphasizes its essential independence of dimensionality.

---

[4] The material of this chapter, in the one-dimensional case, has been discussed by the author with Professor Eberhard Hopf several years ago, and he wishes to thank Professor Hopf for suggestions which have contributed to his present point of view.

The type of chaos which we shall consider is that in which the expression $\mathfrak{F}(\Sigma;\alpha)$ has a distribution in α dependent only on the measure of the set Σ; and in which, if Σ_1 and Σ_2 do not overlap, the distributions of $\mathfrak{F}(\Sigma_1,\alpha)$ and $\mathfrak{F}(\Sigma_2,\alpha)$ are independent, in the sense that if $\phi(x,y)$ is a measurable function, and either side of the equation has a sense,

$$(56)\qquad \int_0^1\int_0^1 \phi(\mathfrak{F}(\Sigma_1;\alpha),\mathfrak{F}(\Sigma_2,\beta))\,d\alpha d\beta = \int_0^1 \phi(\mathfrak{F}(\Sigma_1,\alpha),\mathfrak{F}(\Sigma_2,\alpha))\,d\alpha.$$

We assume a similar independence when n non-overlapping sets $\Sigma_1,\cdots,\Sigma_n$ are concerned. It is by no means intuitively certain that such a type of chaos exists. In establishing its existence, we encounter a difficulty belonging to many branches of the theory of the Lebesgue integral. The fundamental theorem of Lebesgue assures us of the possibility of adding the measures of a denumerable assemblage of measurable sets, to get the measure of their sum, if they do not overlap. Accordingly, behind any effective realization of the theory of Lebesgue integration, there is always a certain denumerable family of sets in the background, such that all measurable sets may be approximated by denumerable combinations of these. This family is not unique, but without the possibility of finding it, there is no Lebesgue theory.

On the other hand, a theory of measure suitable for the description of a chaos must yield the measure of any assemblage of functions arising from a given measurable assemblage by a translational change of origin. This set of assemblages is essentially non-denumerable. Any attempt to introduce the notion of measure in a way which is invariant under translational changes of origin, without the introduction of some more restricted set of measurable sets, which does not possess this invariance, will fail to establish those essential postulates of the Lebesgue integral which deal with denumerable sets of points. There is no way of avoiding the introduction of constructional devices which seem to restrict the invariance of the theory, although once the theory is obtained it may be established in its full invariance.

Accordingly, we shall start our theory of randomness with a division of space, whether of one dimension or of more, into a denumerable assemblage of sub-sets. In one dimension, this division may be that into those intervals whose coördinates are terminating binary numbers, and in more dimensions, into those parallelepipeds with edges parallel to the axes and with terminating binary coördinates for the corner points. We then wish to find a self-consistent distribution-function for the mass in such a region, dependent only on the volume, and independent for non-overlapping regions.

This problem does not admit of a unique solution, although the solution becomes essentially unique if we adjoin suitable auxiliary conditions. Among these conditions, for example, is the hypothesis that the distribution is symmetric, as between positive and negative values, has a finite mean square, and that the measure of the set of α's for which $\mathfrak{F}(S;\alpha) > A$ is a continuous function of A.[5] Without going into such considerations, we shall assume directly that the measure of the set of instances in which the value of $\mathfrak{F}(S;\alpha)$ in a region S of measure M lies between a and $b > a$, is

$$(57) \qquad \frac{1}{\sqrt{2\pi M}}\int_a^b \exp\left(-\frac{u^2}{2M}\right)du.$$

The formula

$$(58) \qquad \frac{1}{\sqrt{2\pi M_1 M_2}}\int_{-\infty}^{\infty} \exp\left(-\frac{u^2}{2M_1}-\frac{(v-u)^2}{2M_2}\right)du = \frac{1}{\sqrt{M_1+M_2}}\exp\left(-\frac{v^2}{2(M_1+M_2)}\right)$$

shows the consistency of this assumption.

The distributions of mass among the sets of our denumerable assemblage may be mapped on the line segment $0 \leq \alpha \leq 1$, in such a way that the measure of the set of instances in which a certain contingency holds will go into a set of values of α of the same measure. This statement needs a certain amount of elucidation. To begin with, the only sets of instances whose measures we know are those determined by

$$(59) \qquad \begin{array}{c} a_1 \leq \mathfrak{F}(S_1;\alpha) \leq b_1 \\ a_2 \leq \mathfrak{F}(S_2;\alpha) \leq b_2 \\ \cdots\cdots\cdots\cdots \\ a_n \leq \mathfrak{F}(S_n;\alpha) \leq b_n; \end{array}$$

where $S_1, S_2, \cdots, S_n$ are to be found among our denumerable set of subdivisions of space. However, once we have established a correspondence between the measures of these specific sets of contingencies and their corresponding sets of values of α, we may use the measure of any measurable set of values of α to define the measure of its corresponding set of contingencies.

The correspondence between sets of contingencies and points on the line $(0,1)$ is made by determining a hierarchy of sets of contingencies

$$(60) \qquad \begin{array}{c} a_1^{(m,n)} \leq \mathfrak{F}(S_1^{(m,n)};\alpha) \leq b_1^{(m,n)} \\ \cdots\cdots\cdots\cdots\cdots \\ a_\nu^{(m,n)} \leq \mathfrak{F}(S_\nu^{(m,n)};\alpha) \leq b_\nu^{(m,n)}. \end{array}$$

---

[5] Cf. the recent investigations of Cramér and P. Lévy.

Let us call such a contingency $C_{m,n}$. If m is fixed, let all the contingencies $C_{m,n}$ $(n = 1, 2, \cdots)$ be mutually exclusive, and let them be finite in number. Let us be able to write

$$C_{m,n} = \sum_{k=1}^{N} C_{m+1,n_k}. \tag{61}$$

If S' is one of our denumerable sets of regions of space, let every $C_{m,n}$ with a sufficiently large index m be included in a class determined by a set of conditions concerning the mass on S' alone, and restricting it to a set of values lying in an interval (c, d), corresponding to an integral of form (57) and of arbitrarily small value. (Here d may be ∞, or c may be $-\infty$.) Let us put $C_{1,1}$ for the entire class of all possible contingencies, and let us represent it by the entire interval $(0, 1)$. Let us assume that $C_{m,n}$ has been mapped into an interval of length corresponding to its probability, in accordance with (57), and let this interval be divided in order of the sequence of their n_k's into intervals corresponding respectively to the component C_{m+1,n_k}'s, and of the same measure. Except for a set of points of zero measure, every point of the segment $(0, 1)$ of α will then be determined uniquely by the sequence of the intervals containing it and corresponding to the contingencies $C_{m,r}$ for successive values of m. This sequence will then determine uniquely (except in a set of cases corresponding to a set of values of α of zero measure) the value of $\mathfrak{F}(S_n; \alpha)$ for every one of our original denumerable set of sets S_n.

So far, everything that we have said has been independent of dimensionality. We now proceed to something belonging specifically to the one-dimensional case. If the original sets S_n are the sets of intervals with binary end-points, of such a form that they may be written in the binary scale

$$(d_1 d_2 \cdots d_k \cdot d_{k+1} \cdots d_l,\ d_1 \cdots d_k \cdot d_{k+1} \cdots d_l + 2^{-l-1}) \tag{62}$$

where $d_1, \cdots, d_l$ are digits which are either 0 or 1, then any interval whatever of length not exceeding $2^{-\mu}$ and lying in (a, b) (where a and b are integers) may be written as the sum of not more than two of the $(b-a)2^{\mu+1}$ intervals of form (62) lying in (a, b) and of length $2^{-\mu-1}$, not more than two of the $(b-a)2^{\mu+2}$ intervals of length $2^{-\mu-2}$, and so on. The probability that the value of $|\mathfrak{F}(S_n; \alpha)|$ should exceed A, or in other words, the measure of the set of α's for which it exceeds A, is

$$\frac{2}{\sqrt{2\pi m(S_n)}} \int_A^\infty \exp\left(-\frac{u^2}{2m(S_n)}\right) du = o\{\exp(-Am(S_n)^{-\frac{1}{2}})\}. \tag{63}$$

Now let us consider the total probability that the value of $|\mathfrak{F}(S_n; \alpha)|$ should exceed $2^{-(\mu+1)(\frac{1}{2}-\epsilon)}$ for any one of the $(b-a)2^{\mu+1}$ intervals of length

$2^{-\mu-1}$, or $2^{-(\mu+2)(\frac{1}{2}-\epsilon)}$ for any one of the $(b-a)2^{\mu+2}$ intervals of length $2^{-\mu-2}$, or so on. This probability can not exceed

$$(64) \qquad \sum_{k=1}^{\infty}(b-a)2^{\mu+k}\,o(\exp(-2^{(\mu+k)\epsilon}))=o(2^{\mu}\exp(-2^{(\mu+1)\epsilon})).$$

On the other hand, the sum of $|\mathfrak{F}(S_n;\alpha)|$ for all the $2+2+\cdots$ intervals must in any other case be equal to or less than

$$(65) \qquad 2\sum_{k=1}^{\infty}2^{-(\mu+k)(\frac{1}{2}-\epsilon)}=O(2^{-\mu(\frac{1}{2}-\epsilon)}).$$

Thus there is a certain sense in which over a finite interval, and except for a set of values of α of arbitrarily small positive measure, the total mass in a sub-interval of length $\leq 2^{-\mu}$ tends uniformly to 0 with $2^{-\mu}$. On this basis, we may extend the functional $\mathfrak{F}(S;\alpha)$ to all intervals S. It is already defined for all intervals with terminating binary end-points. If (c,d) is any interval whatever, let $c_1, c_2, \cdots$ be a sequence of terminating binary numbers approaching c, and let $d_1, d_2, \cdots$ be a similar sequence approaching d. Then except for a fixed set of values of α of arbitrarily small measure,

$$(66) \qquad \lim_{m,n\to\infty}|\mathfrak{F}((c_n,d_n)+\alpha)-\mathfrak{F}((c_m,d_m);\alpha)|=0,$$

and we may put

$$(67) \qquad \mathfrak{F}((c,d);\alpha)=\lim_{m\to\infty}\mathfrak{F}((c_m,d_m);\alpha).$$

Formula (57), and the fact that $\mathfrak{F}(S_1;\alpha)$ and $\mathfrak{F}(S_2;\alpha)$ vary independently for non-overlapping intervals S_1 and S_2, will be left untouched by this extension.

Thus if Σ is an interval and T^{λ} a translation through an amount λ, we can define $\mathfrak{F}(T^{\lambda}\Sigma;\alpha)$, and it will be equally continuous in λ over any finite range of λ except for a set of values of α of arbitrarily small measure. From this it follows at once that it is measurable in λ and α together. Furthermore, we shall have

$$(68) \qquad \begin{array}{l}\text{Measure of set of } \alpha\text{'s for which } \mathfrak{F}(T^{\lambda}\Sigma;\alpha) \text{ belongs to } C = \\ \text{Measure of set of } \alpha\text{'s for which } \mathfrak{F}(\Sigma;\alpha) \text{ belongs to } C.\end{array}$$

Thus $\mathfrak{F}(\Sigma;\alpha)$ is a homogeneous chaos. We shall call it *the pure chaos.*

If $\Phi\{\mathfrak{F}(\Sigma;\alpha)\}$ is a functional dependent on the values of $\mathfrak{F}(\Sigma;\alpha)$ for a finite number of intervals Σ_n, then if λ is so great that none of the intervals Σ_n overlaps any translated interval $T^{\lambda}\Sigma_m$, $\Phi\{\mathfrak{F}(\Sigma;\alpha)\}$ will have a distribution entirely independent of $\Phi\{\mathfrak{F}(T^{\lambda}\Sigma;\alpha)\}$. As every measurable functional may be approached in the L sense by such a functional, we see at once that $\mathfrak{F}(\Sigma;\alpha)$ is metrically transitive.[6]

---

[6] Except that the method of treatment has been adapted to the needs of § 7, the

7. Pure multidimensional chaos. In order to avoid notational complexity, we shall not treat the general multidimensional case explicitly, but shall treat the two-dimensional case by a method which will go over directly to the most general multidimensional case. If our initial sets S_n are the rectangles with terminating binary coördinates for their corners and sides parallel to the axes, and we replace (a, b) by the square with opposite vertices (p, q) and $(p+r,\ q+r)$, an argument of exactly the same sort as that which we have used in the last paragraph will show that except in a set of cases of total probability not exceeding

$$(69) \qquad \text{const.} \sum_{k=1}^{\infty} \sum_{l=1}^{\infty} 2^{\mu+k} 2^{l}\, o\left(\exp\left(-2^{(\mu+k+l)\epsilon}\right)\right) = o\left(2^{\mu} \exp\left(-2^{(\mu+1)\epsilon}\right)\right)$$

the sum of $|\mathfrak{F}(S_n;\alpha)|$ for a denumerable set of binary rectangles with base $\leq 2^{-\mu}$, of the form (62), and adding up to make a vertical interval lying in the square (p, q), $(p+r, q+r)$, must be equal to or less than

$$(65) \qquad 2\sum_{k=1}^{\infty} 2^{-(\mu+k)(\frac{1}{2}-\epsilon)} = O\left(2^{-\mu(\frac{1}{2}-\epsilon)}\right).$$

If we now add this expression up for all the base intervals of type (62) necessary to exhaust a horizontal interval of magnitude not exceeding $2^{-\mu}$, we shall again obtain an expression of the form (65). It hence follows that if we take the total mass on the coördinate rectangles within a given square, this will tend to zero uniformly with their area, except for a set of values of α of arbitrarily small measure. From this point the two-dimensional argument, and indeed the general multidimensional argument, follows exactly the same lines as the one-dimensional argument. It is only necessary to note that if

$$a_n \to a,\ b_n \to b,\ c_n \to c,\ d_n \to d$$

then the rectangles (a_m, b_m), (c_m, d_m) and (a_n, b_n), (c_n, d_n) differ at most by four rectangles of small area.[7]

From this point on, we shall write $\mathfrak{P}(S;\alpha)$ for a pure chaos, whether in one or in more dimensions.

8. Phase averages in a pure chaos. If $f(P)$ is a measurable step-function, the definition of

$$(70) \qquad \int f(P)\, d_P \mathfrak{P}(S;\alpha)$$

---

results of this section have previously been demonstrated by the author. (*Proceedings of the London Mathematical Society*, 2, vol. 22 (1924), pp. 454-467).

[7] Here we represent a rectangle by giving two opposite corners.

is obvious, for it reduces to the finite sum

$$\sum_{1}^{N} f_n \mathfrak{P}(S_n;\alpha) \tag{71}$$

where f_n are the N values assumed by $f(P)$, and S_n respectively are the sets over which these values are assumed. Let us notice that

$$\begin{aligned} \int_0^1 d\alpha \left| \int f(P)\, d_P \mathfrak{P}(S;\alpha) \right|^2 &= \sum_{m=1}^{N} \sum_{n=1}^{N} \int_0^1 d\alpha f_m \bar{f}_n \mathfrak{P}(S_m;\alpha)\, \mathfrak{P}(S_n;\alpha) \\ &= \sum_{1}^{N} |f_n|^2 \int_0^1 (\mathfrak{P}(S_n;\alpha))^2 d\alpha \\ &= \sum_{1}^{N} |f_n|^2 \frac{1}{\sqrt{2\pi m(S_n)}} \int_{-\infty}^{\infty} u^2 \exp\left(-\frac{u^2}{2m(S_n)}\right) du \\ &= \sum_{1}^{N} |f_n|^2 m(S_n) \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} u^2 e^{-(u^2/2)} du \\ &= \sum_{1}^{N} |f_n|^2 m(S_n) = \int |f(P)|^2 dV_P, \end{aligned} \tag{72}$$

the integral being taken over the whole of space. In other words, the transformation from $f(P)$ as a function of P, to $\int f(P)\, d_P \mathfrak{P}(S;\alpha)$ as a function of α, retains distance in Hilbert space.[8] Such a transformation, by virtue of the Riesz-Fischer theorem, may always be extended by making limits in the mean correspond to limits in the mean. Thus both in the one-dimensional and in the many-dimensional case, we may *define*

$$\int f(P)\, d_P \mathfrak{P}(S;\alpha) = \underset{n\to\infty}{\text{l. i. m.}} \int f_n(P)\, d_P \mathfrak{P}(S;\alpha) \tag{73}$$

where $f(P)$ is a function belonging to L^2, and the sequence $f_1(P), f_2(P), \cdots$ is a sequence of step-functions converging in the mean to $f(P)$ over the whole of space. The definition will be unambiguous, except for a set of values of α of zero measure.

If S is any measurable set, we have

$$\begin{aligned} \int_0^1 d\alpha \{\mathfrak{P}(S;\alpha)\}^n &= \frac{1}{\sqrt{2\pi m(S)}} \int_{-\infty}^{\infty} u^n \exp\left(-\frac{u^2}{2m(S)}\right) du \\ &= (m(S))^{n/2} \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} u^n e^{-(u^2/2)} du \\ &\left\{ \begin{array}{l} = 0 \text{ if } n \text{ is odd} \\ = (m(S))^{n/2}(n-1)(n-3)\cdots 1 \text{ if } n \text{ is even.} \end{array} \right. \end{aligned} \tag{74}$$

---

[8] Cf. Paley, Wiener, and Zygmund, *Mathematische Zeitschrift*, vol. 37 (1933), pp. 647-668.

This represents $(m(S))^{n/2}$, multiplied by the number of distinct ways of representing n objects as a set of pairs. Remembering that if $S_1, S_2 \cdots, S_{2n}$ are non-overlapping, their distributions are independent, we see that if the sets $\Sigma_1, \Sigma_2, \cdots, \Sigma_{2n}$ are either totally non-overlapping, or else such that when two overlap, they coincide, we have[9]

$$(75) \qquad \int_0^1 \mathfrak{P}(\Sigma_1;\alpha)\cdots\mathfrak{P}(\Sigma_n;\alpha)\,d\alpha = \Sigma\Pi\int_0^1 \mathfrak{P}(\Sigma_j;\alpha)\,\mathfrak{P}(\Sigma_k;\alpha)\,d\alpha,$$

where the product sign indicates that the $2n$ terms are divided into n sets of pairs, j and k, and that these factors are multiplied together, while the addition is over all the partitions of $1, \cdots, 2n$ into pairs. If $2n$ is replaced by $2n+1$, the integral in (75) of course vanishes.

Since $\mathfrak{P}(S;\alpha)$ is a linear functional of sets of points, and since both sides of (75) are linear with respect to each $\mathfrak{P}(\Sigma_k;\alpha)$ separately, (75) still holds when $\Sigma_1, \Sigma_2, \cdots, \Sigma_{2n}$ can be reduced to sums of sets which either coincide or do not overlap, and hence holds for all measurable sets.

Now let $f(P_1, \cdots, P_n)$ be a measurable step-function: that is, a function taking only a finite set of finite values, each over a set of values $P_1, \cdots, P_n$ which is a product-set of measurable sets in each variable P_k. Clearly we may define

$$(76) \qquad \int\cdots\int f(P_1,\cdots,P_n)\,d_{P_1}\mathfrak{P}(S;\alpha)\cdots d_{P_n}\mathfrak{P}(S;\alpha)$$

in a way quite analogous to that in which we have defined (70), and we shall have

$$(77) \qquad \int_0^1 d\alpha\int\cdots\int f(P_1,\cdots,P_n)\,d_{P_1}\mathfrak{P}(S;\alpha)\cdots d_{P_n}\mathfrak{P}(S;\alpha)$$
$$= \Sigma\int\cdots\int f(P_1,P_1,P_2,P_2,\cdots,P_n,P_n)\,dV_{P_1}\cdots dV_{P_n}$$

where the summation is carried out for all possible divisions of the $2n$ P's into pairs. Similarly in the odd case

$$(78) \qquad \int_0^1 d\alpha\int\cdots\int f(p_1,\cdots,P_{2n+1})\,d_{P_1}\mathfrak{P}(S;\alpha)\cdots d_{P_n}\mathfrak{P}(S;\alpha) = 0.$$

We may apply (77) to give a meaning to

$$(79) \qquad \int_0^1 d\alpha\left|\int\cdots\int f(P_1,\cdots,P_n)\,d_{P_1}\mathfrak{P}(S;\alpha)\cdots d_{P_n}\mathfrak{P}(S;\alpha)\right|^2.$$

If $f(P_1,\cdots,P_n)$ is a measurable step-function, and

---

[9] Cf. Paley, Wiener, and Zygmund, *loc. cit.*, formula (2.05).

$$(80) \qquad |f(P_1, \cdots, P_n)| \leq |f_1(P_1) \cdots f_n(P_n)|\,; \qquad \int |f_k(P)|^2 dV_P \leq A \qquad (k = 1, 2, \cdots)$$

we shall have

$$(81) \qquad \int_0^1 d\alpha \left| \int \cdots \int f(P_1, \cdots P_n)\, d_{P_1}\mathfrak{P}(S;\alpha) \cdots d_{P_n}\mathfrak{P}(S;\alpha) \right|^2 \leq A^n (2n-1)(2n-3)\cdots 1.$$

If now $f(P_1, \cdots, P_n)$ is an integrable function satisfying (80), but not necessarily a step-function, let

$$(82) \quad f(\nu; P_1, \cdots, P_n) = \frac{1}{\nu} \operatorname{sgn} f(P_1, \cdots, P_n)\, [\nu f(P_1, \cdots, P_n) \operatorname{sgn} f(P_1, \cdots, P_n)].$$

Clearly almost everywhere

$$(83) \qquad \int \cdots \int f(\nu; P_1, \cdots, P_n)\, d_{P_1}\mathfrak{P}(S;\alpha) \cdots d_{P_n}\mathfrak{P}(S;\alpha) \leq \prod_1^n \int f_k(P)\, d_P\mathfrak{P}(S;\alpha)$$

and

$$(84) \qquad \overline{\lim_{\mu,\nu\to\infty}} \left| \int \cdots \int f(\mu; P_1, \cdots, P_n)\, d_{P_1}\mathfrak{P}(S;\alpha) \cdots d_{P_n}\mathfrak{P}(S;\alpha) - \int \cdots \int f(\nu; P_1, \cdots, P_n)\, d_{P_1}\mathfrak{P}(S;\alpha)\, d_{P_2}\mathfrak{P}(S;\alpha) \right|$$
$$\leq \prod_1^n \int f_k(P)\, d_P\mathfrak{P}(S;\alpha) \left(\epsilon + \sum_1^n \left\{ \frac{\int_R f_k(P)\, d_P\mathfrak{P}(S;\alpha)}{\int f_k(P)\, d_P\mathfrak{P}(S;\alpha)} \right\} \right)$$

where R represents the exterior of a sphere of arbitrarily large volume. Let it be noted that both the numerator and the denominator of this fraction have Gaussian distributions, but that the mean square value of the numerator is arbitrarily small. Thus except for a set of values of α of arbitrarily small measure, the right side of expression (84) is arbitrarily small, so that we may write

$$(85) \qquad \lim_{\mu,\nu\to\infty} \left\{ \int \cdots \int f(\mu; P_1, \cdots, P_n)\, d_{P_1}\mathfrak{P}(S;\alpha) \cdots d_{P_n}\mathfrak{P}(S;\alpha) - \int \cdots \int f(\nu; P_1, \cdots, P_n)\, d_{P_1}\mathfrak{P}(S;\alpha) \cdots d_{P_n}\mathfrak{P}(S;\alpha) \right\} = 0.$$

Thus by dominated convergence,

$$(86) \qquad \lim_{\mu\to\infty} \int \cdots \int f(\mu; P_1, \cdots, P_n)\, d_{P_1}\mathfrak{P}(S;\alpha) \cdots d_{P_n}\mathfrak{P}(S;\alpha)$$

exists for almost all values of α, and we may write it by definition

$$(87) \qquad \int \cdots \int f(P_1, \cdots, P_n)\, d_{P_1}\mathfrak{P}(S;\alpha) \cdots d_{P_n}\mathfrak{P}(S;\alpha).$$

This will clearly be unique, except for a set of values of α of zero measure. There will then be no difficulty in checking (77), (78), and (81).

9. Forms of chaos derivable from a pure chaos. Let us assume that $f(P)$ belongs to L^2, or that $f(P_1, \cdots, P_n)$ is a measurable function satisfying (80). Let us write $\overset{\frown}{PQ}$ for the vector in n-space connecting the points P and Q. Then the function

$$(88) \qquad \int \cdots \int f(\overset{\frown}{PP_1}, \cdots, \overset{\frown}{PP_n})\, d_{P_1}\mathfrak{P}(S;\alpha) \cdots d_{P_n}\mathfrak{P}(S;\alpha) = F(P;\alpha)$$

is a metrically transitive differentiable chaos. This results from the fact that $\mathfrak{P}(S;\alpha)$ is a metrically transitive chaos, and that a translation of P generates a similar translation of all the points P_k. The sum of a finite number of functions of the type (88) is also a metrically transitive differentiable chaos. To show that $F(P;\alpha)$ is measurable in P and α simultaneously, we merely repeat the argument of (83)-(86) with both P and α as variables.

We shall call a chaos such as (88) a *polynomial chaos homogeneously of the n-th degree,* and a sum of such chaoses a *polynomial chaos* of the degree of its highest term. In this connection, we shall treat a constant as a chaos homogeneously of degree zero.

By the multidimensional ergodic theorem, if Φ is a functional such that

$$(89) \qquad \int_0^1 |\Phi(F(P;\alpha))| \log^+ |\Phi(F(P;\alpha))|\, d\alpha < \infty$$

we shall have

$$(90) \qquad \lim_{r\to\infty} \frac{1}{V(r)} \int_R \Phi(F(P;\alpha))\, dV_P = \int_0^1 \Phi(F(P;\alpha))\, d\alpha$$

for almost all values of α. Since the distribution of $F(P;\alpha)$ is dominated by the product of a finite number of independent Gaussian distributions, we even have

$$(91) \qquad \int_0^1 |F(P;\alpha)|^n d\alpha < \infty$$

for all positive integral values of n. In a wide class of cases this enables us to establish a relation of the type of (89).

In formula (90), we have an algorithm for the computation of the right-hand side. For example, if

$$(92)\qquad F(P;\alpha)=\int f(\widehat{PP_1})\,d_{P_1}\mathfrak{P}(S;\alpha),$$

and $P+Q$ is the vector sum of P and Q, we have for almost all α,

$$(93)\qquad \lim_{r\to\infty}\frac{1}{V(r)}\int_R F(P+Q;\alpha)\bar{F}(Q;\alpha)\,dV_Q=\int f(P+Q)\bar{f}(Q)\,dV_Q,$$

the integral being taken over the whole of space; if

$$(94)\qquad F(P;\alpha)=\int\cdots\int f(\widehat{PP_1},\widehat{PP_2})\,d_{P_1}\mathfrak{P}(S;\alpha)\,d_{P_2}\mathfrak{P}(S;\alpha),$$

we have almost always

$$\begin{aligned}(95)\qquad &\lim_{r\to\infty}\frac{1}{V(r)}\int_R F(P+Q;\alpha)\bar{F}(Q;\alpha)\,dV_Q\\ &=\left|\int f(Q,Q)\,dV_Q\right|^2+\int\int f(P+Q,P+M)\bar{f}(Q,M)\,dV_Q\,dV_M\\ &\quad+\int\int f(P+Q,P+M)\bar{f}(M,Q)\,dV_Q\,dV_M;\end{aligned}$$

and if

$$(96)\quad F(P;\alpha)=\int\int\int f(\widehat{PP_1},\widehat{PP_2},\widehat{PP_3})\,d_{P_1}\mathfrak{P}(S;\alpha)\,d_{P_2}\mathfrak{P}(S;\alpha)\,d_{P_3}\mathfrak{P}(S;\alpha),$$

we have almost everywhere

$$\begin{aligned}(97)\quad &\lim_{r\to\infty}\frac{1}{V(r)}\int_R F(P+Q;\alpha)\bar{F}(Q;\alpha)\,dV_Q\\ &=\int\int\int\{f(Q,Q,P+M)\bar{f}(M,S,S)\\ &\quad+f(Q,Q,P+M)\bar{f}(S,M,S)+f(Q,Q,P+M)\bar{f}(S,S,M)\\ &\quad+f(Q,P+M,Q)\bar{f}(M,S,S)+f(Q,P+M,Q)\bar{f}(S,M,S)\\ &\quad+f(Q,P+M,Q)\bar{f}(S,S,M)+f(P+M,Q,Q)\bar{f}(M,S,S)\\ &\quad+f(P+M,Q,Q)\bar{f}(S,M,S)+f(P+M,Q,Q)\bar{f}(S,S,M)\\ &\quad+f(P+Q,P+M,P+S)\bar{f}(Q,M,S)\\ &\quad+f(P+Q,P+M,P+S)\bar{f}(Q,S,M)\\ &\quad+f(P+Q,P+M,P+S)\bar{f}(M,Q,S)\\ &\quad+f(P+Q,P+M,P+S)\bar{f}(M,S,Q)\\ &\quad+f(P+Q,P+M,P+S)\bar{f}(S,Q,M)\\ &\quad+f(P+Q,P+M,P+S)\bar{f}(S,M,Q)\}\,dV_Q\,dV_M\,dV_S.\end{aligned}$$

We have similar results in the non-homogeneous case. Thus if

$$\begin{aligned}(98)\qquad F(P;\alpha)&=A+\int f(\widehat{PP_1})\,d_{P_1}\mathfrak{P}(S;\alpha)\\ &\quad+\int\int g(\widehat{PP_1},\widehat{PP_2})\,d_{P_1}\mathfrak{P}(S;\alpha)\,d_{P_2}\mathfrak{P}(S;\alpha),\end{aligned}$$

we have almost everywhere

$$(99)\qquad \begin{aligned}\lim_{r\to\infty}\frac{1}{V(r)}\int_R F(P+Q;\alpha)\bar{F}(Q;\alpha)\,dV_Q &= A\int\bar{g}(Q,Q)\,dV_Q+\bar{A}\int g(Q,Q)\,dV_Q\\ &+\int f(P+Q)\bar{f}(Q)\,dV_Q+\left|\int g(Q,Q)\,dV_Q\right|^2\\ &+\int\int g(P+Q,P+M)\bar{g}(Q,M)\,dV_Q\,dV_M\\ &+\int\int g(P+Q,P+M)\bar{g}(M,Q)\,dV_Q\,dV_M.\end{aligned}$$

10. Chaos theory and spectra.[10] The function

$$(100)\qquad \lim_{r\to\infty}\frac{1}{V(r)}\int_R F(P+Q)\bar{F}(Q)\,dV_Q=G(P)$$

occupies a central position in the theory of harmonic analysis. If it exists and is continuous for every value of P, the function $F(P)$ is said to have an n-dimensional spectrum. To define this spectrum, we put

$$(101)\qquad F_r(P)=\begin{cases}F(P) & \text{on } R;\\ 0 & \text{elsewhere.}\end{cases}$$

It is then easy to show by an argument involving considerations like those of (49) that if

$$(102)\qquad G_r(P)=\frac{1}{V(r)}\int_\infty F_r(P+Q)\bar{F}_r(Q)\,dV_Q,$$

the integral being taken over the whole of space, then we have

$$(103)\qquad G(P)=\lim_{r\to\infty}G_r(P).$$

Since, if O is the point with zero coördinates, by the Schwarz inequality,

$$(104)\qquad |G(P)|\leq G(O),$$

the limit in (103) is approached boundedly.

If now we put

$$(105)\qquad \phi_r(U)=(2\pi)^{-(n/2)}V(r)^{-\frac{1}{2}}\operatorname*{l.i.m.}_{s\to\infty}\int_S F_r(P)e^{iU\cdot P}\,dV_P$$

where S is the interior of a sphere of radius s about the origin, the n-fold Parseval theorem will give us

---

[10] Cf. N. Wiener, "Generalized harmonic analysis," *Acta Mathematica*, vol. 55 (1930).

$$|\phi_r(U)|^2 = (2\pi)^{-n}\int_\infty G_r(P)e^{iU\cdot P}dV_P. \tag{106}$$

If $M(U)$ is a function with an absolutely integrable Fourier transform, we shall have

$$\int_{-\infty}^{\infty}|\phi_r(U)|^2 M(U)\,dV_U = (2\pi)^{-n}\int_\infty G_r(P)\,dV_P\int_\infty M(U)e^{iU\cdot P}dV_U, \tag{107}$$

and hence

$$\lim_{r\to\infty}\int_{-\infty}^{\infty}|\phi_r(U)|^2 M(U)\,dV_U = (2\pi)^{-n}\int G(P)\,dV_P\int_\infty M(U)e^{iU\cdot P}dV_U \tag{108}$$

which will always exist. Let us put

$$\mathfrak{M}\{M(U)\} = (2\pi)^{-n}\int_\infty G(P)\,dV_P\int_\infty M(U)e^{iU\cdot P}dV_U. \tag{109}$$

If S is any set of points of finite measure, and $S(P)$ is its characteristic function, let us put

$$\overline{\mathfrak{M}}(S) = \underset{M(U)\geqq S(U)}{\text{l. u. b.}}\ \mathfrak{M}(M(U)), \tag{110}$$

and

$$\underline{\mathfrak{M}}(S) = \underset{M(U)\leqq S(U)}{\text{g. l. b.}}\ \mathfrak{M}(M(U)). \tag{111}$$

If $\underline{\mathfrak{M}}(S)$ and $\overline{\mathfrak{M}}(S)$ have the same value, we shall write it $\mathfrak{M}(S)$, and shall call it the *spectral mass* of F on S. It will be a non-negative additive set-function of S, and may be regarded as determining the spectrum of F.

If $f(P_1, \cdots, P_n)$ satisfies (80) and $F(P;\alpha)$ is defined as in (88), we know that for any given P,

$$G(P;\alpha) = \lim_{r\to\infty}\frac{1}{V(r)}\int_R F(P+Q;\alpha)\bar{F}(Q;\alpha) = \int_0^1 F(P,\beta)\bar{F}(0,\beta)\,d\beta \tag{112}$$

for almost all values of α. This alone is not enough to assure that $F(\mathrm{P},\alpha)$ has a spectrum for almost all values of α, as the sum of a non-denumerable set of sets of zero measure is not necessarily of zero measure. On the other hand, except for a set of values of α of zero measure, $G(P,\alpha)$ exists for all points P with rational coördinates.

We may even extend this result, and assert that if

$$F_\theta(P;\alpha) = \frac{1}{V(\theta)}\int_{\text{length of } \widehat{PS}\leqq\theta} F(S;\alpha)\,dV_S \tag{113}$$

and

$$G_\theta(P;\alpha) = \lim_{r\to\infty}\frac{1}{V(r)}\int_R F_\theta(P+Q;\alpha)\bar{F}_\theta(Q;\alpha)\,dV_Q, \tag{114}$$

then except for a set of values of α of zero measure, $G_\theta(P;\alpha)$ exists for all points P with rational coördinates and all rational parameters θ, and it is easily proved that for almost all values of α, as θ tends to 0 through rational values,

$$\lim_{\theta\to 0}\lim_{r\to\infty}\frac{1}{V(r)}\int_R |F_\theta(Q;\alpha)-F(Q;\alpha)|^2\,dV_Q=0. \tag{115}$$

Now, by the Schwarz inequality,

$$\begin{aligned}
\tag{116} &\left|\frac{1}{V(r)}\int_R F_\theta(P+Q;\alpha)\bar F_\theta(Q;\alpha)\,dV_Q\right.\\
&\left.\quad-\frac{1}{V(r)}\int_R F_\theta(P_1+Q;\alpha)\bar F_\theta(Q;\alpha)\,dV_Q\right|\\
&\leq\left\{G_\theta(O;\alpha)\left(\frac{1}{V(r)}\int_R |F_\theta(P+Q;\alpha)-F_\theta(P_1+Q;\alpha)|^2\,dV_Q\right)\right\}^{\frac12}\\
&\leq\left\{G_\theta(O;\alpha)\left(\frac{1}{V(r)}\int_R dV_Q\left(\frac{1}{V(\theta)}\left[\int_{|\widehat{P+Q,S}|\leq\theta}-\int_{|\widehat{P+Q,S}|\leq\theta}\right]|F(S;\alpha|^2\,dV_S\right)\right.\right.\\
&\qquad\left.\left.\times\left(\frac{1}{V(\theta)}\left[\int_{|\widehat{P+Q,S}|\leq\theta}-\int_{|\widehat{P_1+Q,S}|\leq\theta}\right]dV_S\right)\right\}^{\frac12}\\
&\leq G_\theta(O;\alpha)\,O(|\widehat{PP_1}|^{\frac12}).
\end{aligned}$$

It thus follows that if (114) exists for a given θ and all P's with rational coördinates, it exists for that θ and all real P's whatever. We may readily show that

$$G_\theta(O;\alpha)\leq G(O;\alpha). \tag{117}$$

By another use of the Schwarz inequality,

$$\begin{aligned}
\tag{118} &\left|\frac{1}{V(r)}\int_R F_\theta(P+Q;\alpha)\bar F_\theta(Q;\alpha)\,dV_Q\right.\\
&\qquad\left.-\frac{1}{V(r)}\int_R F(P+Q;\alpha)\bar F(Q;\alpha)\,dV_Q\right|\\
&\quad\leq\frac{1}{V(r)}\int_R |F_\theta(P+Q;\alpha)-F(P-Q;\alpha)|\;|F_\theta(Q;\alpha)|\,dV_Q\\
&\qquad+\frac{1}{V(r)}\int_R |F(P+Q;\alpha)|\;|F_\theta(Q;\alpha)-F(Q;\alpha)|\,dV_Q\\
&\quad\leq\{G(O;\alpha)\frac{1}{V(r)}\int_R |F_\theta(P+Q;\alpha)-F(P+Q;\alpha)|^2\,dV_Q\}^{\frac12}\\
&\qquad+\{G(O;\alpha)\frac{1}{V(r)}\int_R |F_\theta(Q;\alpha)-F(Q;\alpha)|^2\,dV_Q\}^{\frac12}.
\end{aligned}$$

Combining (115) and (118), we see that except for a set of values of α of zero measure, we have for all P,

$$G(P;\alpha)=\lim_{\theta\to 0}G_\theta(P;\alpha). \tag{119}$$

We thus have an adequate basis for spectrum theory. This will extend, not merely to functions $F(P,\alpha)$ defined as in (88), but to finite sums of such functions. It will even extend to the case of any differentiable chaos $F(P;\alpha)$, for which $F(P+Q;\alpha)\bar{F}(Q;\alpha)$ is an integrable function of α, and for which (115) holds. For a metrically transitive chaos, this latter will be true if

$$(120)\qquad \lim_{\theta\to 0}\int_0^1 |F_\theta(P;\alpha)-F(P;\alpha)|^2\,d\alpha=0.$$

Under this assumption, we have proved that $F(P)$ has a spectrum, and the same spectrum, for all values of α.

This enables us to answer a question which has been put several times, as to whether there is any relation between the spectrum of a chaos and the distribution of its values. There is no unique relation of the sort. The function

$$(121)\qquad \int g(P+Q)\bar{g}(Q)\,dV_Q,$$

where g belongs to L^2, may be so chosen as to represent any Fourier transform of a positive function of L, and if $f(P,Q,M)$ is a bounded step-function, the right-hand side of (97) will clearly be the Fourier transform of a positive function of L. In particular, let $f(P_1,P_2,P_3)=f(P_1)f(P_2)f(P_3)$,

$$(122)\qquad F_1(P;\alpha)=\int f_1(\widehat{PP_1})\,d_{P_1}\mathfrak{P}(S;\alpha)$$

and choose $f_1(Q)$ in such a way that

$$(123)\qquad \int f_1(P+Q)\bar{f}_1(Q)\,dV_Q=\text{right-hand side of (97).}$$

Then

$$(124)\quad \int_0^1 (F_1(P;\alpha))^{2n}d\alpha=\left(\int_\infty |f_1(P)|^2\,dV_P\right)^n(2n-1)(2n-3)\cdots 1$$

and if $F(P,\alpha)$ is defined as in (96),

$$(125)\quad \int_0^1 (F(P;\alpha))^{2n}d\alpha=\left(\int |f(P)|^2dV_P\right)^{3n}(6n-1)(6n-3)\cdots 1)$$

so that for all but at most one value of n,

$$(126)\qquad \int_0^1 (F_1(P;\alpha)^{2n}d\alpha\neq\int_0^1 (F(P;\alpha))^{2n}d\alpha$$

and we obtain in F and F_1 two chaoses with identical spectra but different distribution functions. On the other hand, if

$$(127)\qquad \int_\infty |f_1(P)|^2\,dV_P=\int_\infty |f_2(P)|^2\,dV_P,$$

the chaoses

$$\int f_1(\widehat{PP_1})\,d_{P_1}\mathfrak{P}(S;\alpha) \tag{128}$$

and

$$\int f_2(\widehat{PP_1})\,d_{P_1}\mathfrak{P}(S;\alpha) \tag{129}$$

will have the same distribution functions, but may have very different spectra.

11. The discrete chaos.[11] Let us now divide the whole of Euclidean n-space dichotomously into sets $S_{m,n}$, such that every two sets S_{m_1,n_1} and S_{m_2,n_2} have the same measure, and that each $S_{m,n}$ is made up of exactly two non-overlapping sets $S_{m+1,k}$. Let us divide all these sets into two categories, "occupied," and "empty." Let us require that the probability that a set be empty depend only on its measure, and that the probability that two non-overlapping sets be empty be the product of the probabilities that each be empty. Let us assume that both empty and occupied sets exist. Let every set contained in an empty set be empty, while if a set be occupied, let at least one-half always be occupied. We thus get an infinite class of schedules of emptiness and occupiedness, and methods analogous to those of paragraph 6 may be used to map the class of these schedules in an almost everywhere one-one way on the line $(0,1)$ of the variable α, in such a way that the set of schedules for which a given finite number of regions are empty or occupied will have a probability equal to the measure of the corresponding set of values of α.

By the independence assumption, the probability that a given set $S_{m,n}$ be empty must be of the form $e^{-Am(S_{m,n})}$. If $S_{m,n}$ is divided into the 2^ν intervals $S_{m+\nu,n_1}, \cdots, S_{m+\nu,n_{2^\nu}}$ at the ν-th stage of sub-division, the probability that just one is occupied and the rest are empty is

$$2^\nu(1-\exp(-Am(S_{m,n})/2^\nu)\exp\left(-\frac{2^\nu-1}{2^\nu}Am(S_{m,n})\right). \tag{130}$$

This contingency at the $\nu+1$-st stage is a sub-case of this contingency at the ν-th stage. If we interpret probability to mean the same thing as the measure of the corresponding set of α's, then by monotone convergence, the probability that at every stage, all but one of the subdivisions of $S_{m,n}$ are empty, while the remaining one is occupied, will be the limit of (130), or

$$Am(S_{m,n})\exp(-Am(S_{m,n})). \tag{131}$$

---

[11] The ideas of this paragraph are related to discussions the author has had with Professor von Neumann, and the main theorem is equivalent to one enunciated by the latter.

Such a series of stages of subdivision will have as its occupied regions exactly those which contain a given point.

The probability that the occupied regions are exactly those which contain two points is the probability that each half of $S_{m,n}$ contain exactly one point, plus the probability that one-half is empty, and that in the occupied half, each quarter will contain exactly one point, plus and so on. This will be

$$(132) \quad \left\{ \frac{Am(S_{m,n})}{2} \exp\left(-\frac{Am(S_{m,n})}{2}\right) \right\}^2 + 2\exp\left(-\frac{Am(S_{m,n})}{2}\right) \left\{ \frac{Am(S_{m,n})}{2} \exp\left(-\frac{Am(S_{m,n})}{4}\right)^2 + \cdots = \exp(-Am(S_{m,n})m(S_{m,n}))^2(\tfrac{1}{4} + \tfrac{1}{8} + \cdots) = \frac{(Am(S_{m,n}))^2}{2} \exp(-Am(S_{m,n})).$$

If the probability that the occupied regions are exactly those containing $k-1$ points is

$$\frac{(Am(S_{m,n}))^{k-1}}{(k-1)!} \exp(-Am(S_{m,n}))$$

then a similar argument will show that the probability that the occupied regions are exactly those containing k points will be

$$(133) \quad \sum_{j=1}^{k-1} \frac{1}{j!} \frac{1}{(k-j)!} (Am(S_{m,n}))^k \exp(-Am(S_{m,n})) \left(1 + \frac{1}{2^{k-1}} + \frac{1}{4^{k-1}} + \cdots\right) = \frac{1}{k!}(2^k - 2)\left(\frac{1}{1 - \frac{1}{2^{k-1}}}\right)(Am(S_{m,n}))^k \exp(-Am(S_{m,n})) = \frac{1}{k!}(Am(S_{m,n}))^k \exp(-Am(S_{m,n})).$$

Thus by mathematical induction, the probability that the occupied regions are exactly those containing k points will be

$$\frac{1}{k!}(Am(S_{m,n}))^k \exp(-Am(S_{m,n}))$$

and the sum of this for all values of k will be

$$(134) \quad \sum_0^\infty \frac{1}{k!}(Am(S_{m,n}))^k \exp(-Am(S_{m,n})) = 1.$$

In other words, except for a set of contingencies of probability zero, the occupied regions will be exactly those containing a given finite number of points.

We may proceed at once from the fact that the probability that a set S_1 contains exactly k points is

$$\frac{1}{k!}(Am(S_1))^k e^{-Am(S_1)}$$

while the probability that the non-overlapping set S_2 contains exactly k points is

$$\frac{1}{k!}(Am(S_2))^k e^{-Am(S_2)}$$

to the fact that the probability that the set S_1+S_2 contains exactly k points is

$$\begin{aligned}(135)\qquad &\sum_0^k \frac{1}{j!}\frac{1}{(k-j)!}(Am(S_1))^j(Am(S_2))^{k-j}e^{-Am(S_1+S_2)}\\ &=\frac{1}{k!}(Am(S_1+S_2))^k e^{-Am(S_1+S_2)}.\end{aligned}$$

From this, by monotone convergence, it follows at once that the probability that any set S which is the sum of a denumerable set of our fundamental regions $S_{m,n}$ should contain exactly k points is

$$\frac{1}{k!}(Am(S))^k e^{-Am(S)}.$$

It is then easy to prove this for all measurable sets S.

We are now in a position to prove that the additive functional $\mathfrak{D}(S;\alpha)$, consisting in the number of points in the region S on the basis of the schedule corresponding to α, is a homogeneous metrically transitive chaos. The rôle which continuity filled in paragraphs 6 and 7, of allowing us to show that $\mathfrak{F}_{y_1,\ldots,y_n}(S;\alpha)$ was measurable in $y_1,\cdots,y_n$ and α, is now filled by the fact that the probability that any of the points in a region lie within a very small distance of the boundary, is for any Jordan region the probability that a small region be occupied, and is small. The metric transitivity of the chaos results as before from the independence of the distribution in non-overlapping regions.

The discrete or Poisson chaos which we have thus defined is the chaos of an infinite random shot pattern, or the chaos of the gas molecules in a perfect gas in statistical equilibrium according to the old Maxwell statistical mechanics. It also has important applications to the study of polycrystalline aggregates, and to similar physical problems.

Two important formulae are

$$(136)\qquad \int_0^1 \mathfrak{D}(S;\alpha)\,d\alpha = e^{-Am(S)}\sum_1^\infty \frac{k}{k!}(Am(S))^k = Am(S),$$

and

$$(137)\quad \int_0^1 (\mathcal{D}(S;\alpha))^2 d\alpha = e^{-Am(S)} \sum_1^\infty \frac{k^2}{k!} (Am(S))^k = (Am(S))^2 + Am(S).$$

Let it be noted that if we define

$$(138)\qquad \int f(P)\, d_P \mathcal{D}(S;\alpha)$$

for a measurable step-function $f(P)$ as in (70), by

$$(139)\qquad \sum_1^N f_n \mathcal{D}(S_n;\alpha),$$

(72) is replaced by

$$\begin{aligned}(140)\quad \int_0^1 d\alpha \,\Big| \int f(P)\, d_P\mathcal{D}(S;\alpha) - A\int_\infty f(Q)\, dV_Q \Big|^2 &= \sum_{m=1}^N \sum_{n=1}^N \int_0^1 d\alpha\, f_m \bar{f}_n \mathcal{D}(S_m;\alpha)\mathcal{D}(S_n;\alpha) \\ &\quad - 2\mathcal{R}\{\bar{A}\sum_{m=1}^\infty \int_0^1 d\alpha\, f_m \mathcal{D}(S_m;\alpha) \int_\infty f(Q)\, dV_Q + \Big| A\int_\infty f(Q)\, dV_Q\Big|^2\} \\ &= \sum_{m=1}^N |f_m|^2 Am(S_m) \\ &= A\int_\infty |f(P)|^2\, dV_P.\end{aligned}$$

Thus the transformation from $f(P)$ as a function of P, to

$$(141)\qquad \int f(P)\, d_P\mathcal{D}(S;\alpha) - A\int_\infty f(Q)\, dV_Q$$

as a function of α, retains distance in Hilbert space, apart from a constant factor, and if $f(P)$ belongs to L and L^2 simultaneously, and $\{f_n(P)\}$ is a sequence of step-functions converging in the mean both in the L sense and in the L^2 sense to $f(P)$, we may *define*

$$\begin{aligned}(142)\quad \int f(P)\, d_P\mathcal{D}(S;\alpha) &= A\int_\infty f(Q)\, dV_Q \\ &\quad + \operatorname*{l.i.m.}_{n\to\infty} \Big(\int f_n(P)\, d_P\mathcal{D}(S;\alpha) - A\int_\infty f_n(Q)\, dV_Q\Big).\end{aligned}$$

As in the case of (73), this definition is substantially unique. We may prove the analogue of (93) in exactly the same way as (93) itself, and shall obtain

$$\begin{aligned}(143)\quad \lim_{r\to\infty} \frac{1}{V(r)} \int_R \Big\{ \int_\infty (f(P+\widehat{Q)M};\alpha)\, d_M\mathcal{D}(S;\alpha) - A\int_\infty f(M)\, dV_M \Big\} \\ \times \Big\{ \int_\infty \bar{f}(\widehat{QM};\alpha)\, d_M\mathcal{D}(S;\alpha) - A\int_\infty \bar{f}(M)\, dV_M \Big\}\, dV_Q \\ = A^2 \int_\infty f(P+Q)\bar{f}(Q)\, dV_Q.\end{aligned}$$

As we may see by appealing to the theory of spectra, one interpretation of this in the one-dimensional case is the following: *If a linear resonator be set into motion by a haphazard series of impulses forming a Poisson chaos, the effect, apart from that of a constant uniform stream of impulses, will have the same power spectrum as the energy spectrum of the response of the resonator to a single impulse.*

12. The weak approximation theorem for the polynomial chaos. We wish to show that the chaoses of paragraph 9 are in some sense everywhere dense in the class of all metrically transitive homogeneous chaoses. We shall show that if $\mathfrak{F}(S;\alpha)$ is any homogeneous chaos in n dimensions, there is a sequence $\mathfrak{F}_k(S;\alpha)$ of polynomial chaoses as defined in paragraph 9, such that if $S_1,\cdots,S_\nu$ is any finite assemblage of bounded measurable sets in n-space selected from among a denumerable set, and

$$\int_0^1 |\mathfrak{F}(S_\lambda;\alpha)|^\mu \, d\alpha < \infty \qquad (\lambda = 1, 2, \cdots, \nu) \tag{144}$$

is finite, then

$$\int_0^1 \mathfrak{F}(S_1;\alpha)\cdots\mathfrak{F}(S_\nu;\alpha)\,d\alpha = \lim_{n\to\infty}\int_0^1 \mathfrak{F}_n(S_1;\alpha)\cdots\mathfrak{F}_n(S_\nu;\alpha)\,d\alpha. \tag{145}$$

We first make use of the fact that if the probability that a quantity u be greater in absolute value than A, be less than

$$\frac{2}{\sqrt{2\pi B}}\int_A^\infty e^{-(u^2/2B)}\,du, \tag{146}$$

then if $\psi(u)$ is any even measurable function bounded over $(-\infty,\infty)$, we may find a polynomial $\psi_\epsilon(u)$, such that the mean value of

$$|\psi(u)-\psi_\epsilon(u)|^n \tag{147}$$

which will be

$$\frac{1}{\sqrt{2\pi B}}\int_{-\infty}^\infty |\psi(u)-\psi_\epsilon(u)|^n e^{-(u^2/2B)}\,du, \tag{148}$$

is less than ϵ. Since it is well known that if $\phi(u)$ is a continuous function vanishing outside a finite interval, and

$$\sum_1^\infty A_n H_n(u)\,e^{-(u^2/2)} \tag{149}$$

is the series for $\phi(u)$ in Hermite functions, then we have uniformly

$$\phi(u) = \lim_{t\to 1-0} \sum_{1}^{\infty} A_n t^n H_n(u) e^{-(u^2/2)}, \tag{150}$$

to establish the existence of $\psi_\epsilon(u)$, we need only prove it in the case in which

$$\psi(u) = u^k e^{-Cu^2} \tag{151}$$

for an arbitrarily small value of C: as for example for $C = 1/4n_1B$. We shall then have

$$\left| \psi(u) - u^k \sum_{0}^{N} \frac{(cu^2)^k}{k!} \right| \leq |u|^k \sum_{0}^{\infty} \frac{(cu^2)^k}{k!} = |u|^k e^{u^2/4n_1B}, \tag{152}$$

so that by dominated convergence, and if we take N large enough, we may make

$$\frac{1}{\sqrt{2\pi B}} \int_{-\infty}^{\infty} |\psi(u) - \psi_\epsilon(u)|^n e^{-(u^2/2B)} du < \epsilon \qquad (n \leq n_1). \tag{153}$$

Now let

$$\psi_K(u) = \begin{cases} 0 & (|u| < K); \\ 1 & (|u| \geq K); \end{cases} \tag{154}$$

and let us put

$$\mathcal{G}(P;\alpha) = \frac{1}{V(r)} \mathcal{P}(S;\alpha) \quad (S = \text{interior of } |\widehat{QP}| \leq r). \tag{155}$$

The chaos

$$\mathcal{L}(P;\alpha) = \psi_K(\mathcal{G}(P;\alpha)) \tag{156}$$

may then be approximated by polynomial chaoses in such a way as to approximate simultaneously to all polynomials in $\mathcal{L}(P;\alpha)$ by corresponding polynomials in the approximating chaoses. Since the distribution of the values of $\mathcal{G}(P;\alpha)$ will be Gaussian, with a root mean square value proportional to a power of r, and $\mathcal{G}(P;\alpha)$ will be independent in spheres of radius η about two points P_1 and P_2 more remote from each other than $2r + 2\eta$, it follows that if we take K to be large enough, we may make the probability that $\mathcal{L}(P;\alpha)$ differs from 0 between two spheres of radii respectively $r + \eta$ and H about a given point where it differs from 0, as small as we wish.

We now form the new chaos

$$\int_{|\widehat{PQ}| < x} \mathcal{L}(Q;\alpha)\, dV_Q, \tag{157}$$

which we may also approximate, with all its polynomial functionals, by a sequence of polynomial chaoses. The use of polynomial approximations tending boundedly to a step function over a finite range will show us that this is also true of the chaos determined by

$$(158)\qquad \psi_\gamma\left(\int_{|\widehat{PQ}|<x} \mathcal{L}(Q;\alpha)\,dV_Q\right) = \mathfrak{M}(P;\alpha).$$

By a proper choice of the parameters, this can be made to have arbitrarily nearly all its mass uniformly distributed over regions arbitrarily near to arbitrarily small spheres, all arbitrarily remote from one another, except in an arbitrarily small fraction of the cases. We then form

$$(159)\qquad \frac{1}{(2\pi k)^{n/2}}\int_\infty (\mathfrak{M}(Q;\alpha)+\delta)\exp\left(-\frac{|\widehat{PQ}|^2}{2k}\right)dV_Q = \mathfrak{N}(P;\alpha)$$

where δ is taken to be very small. This chaos again, as far as all its polynomial functionals are concerned, will be approximable by polynomial chaoses. Since it is bounded away from 0 and ∞, and since over such a range the function $1/x$ may be approximated uniformly by polynomials, it follows that in our sense,

$$(160)\qquad 1/\mathfrak{N}(P;\alpha)$$

is approximable by polynomial chaoses.

If $\varpi(P)$ is any measurable function for which arbitrarily high moments are always finite, it is easy to show that

$$(161)\qquad \frac{1}{(2\pi k)^{n/2}}\int_\infty \varpi(Q)(\mathfrak{M}(Q;\alpha)+\delta)\exp\left(-\frac{|\widehat{PQ}|^2}{2k}\right)dV_Q = \mathfrak{W}(P;\alpha)$$

is approximable by polynomial chaoses. Multiplying expressions (160) and (161), it follows that

$$(162)\qquad \mathfrak{W}(P;\alpha)/\mathfrak{N}(P;\alpha) = \mathfrak{U}(P;\alpha)$$

is approximable by polynomial chaoses.

If A is a large enough constant, depending on the choice of the constant ϵ we have

$$(163)\qquad \begin{aligned}&\frac{1}{(2\pi k)^{n/2}}\int_{|P|>A}\exp\left(-\frac{|P|^2}{2k}\right)dV_P\\ &\qquad = \int_A^\infty x^n e^{-(x^2/2k)}\,dx \Big/ \int_0^\infty x^n e^{-(x^2/2k)}\,dx\\ &\qquad < \frac{1}{(2\pi k)^{n/2}}\exp\left(-\frac{(A-\epsilon)^2}{2k}\right).\end{aligned}$$

Thus by the proper choice of the parameters of $\mathfrak{M}(P;\alpha)$, if we take k small enough and then δ small enough, the chaos (162) will consist as nearly as we wish, from the distribution standpoint, of an infinite assemblage of convex cells of great minimum dimension, in each of which the function $\varpi(P)$ is repeated, with the origin moved to some point remote from the boundary.

Now let $\mathfrak{F}(S;\alpha)$ be a metrically transitive homogeneous chaos. Let us form

$$\mathfrak{F}(r;S;\alpha)=\frac{1}{V(r)}\int_R \mathfrak{F}_{x_1,\ldots,x_n}(S;\alpha)\,dx_1\cdots dx_n. \tag{164}$$

Clearly by the fundamental theorem of the calculus, over any finite region in $(x_1,\cdots,x_n)$, we shall have for almost all points and almost all values of α,

$$\mathfrak{F}(S;\alpha)=\lim_{r\to 0}\mathfrak{F}(r;S;\alpha); \tag{165}$$

and if (144) holds, it is easy to show that

$$\int_0^1 |\,\mathfrak{F}(r;S;\alpha)|^n d\alpha < \text{const.} \tag{166}$$

From this it follows that

$$\lim_{r\to 0}\int_0^1 |\,\mathfrak{F}(r;S;\alpha)-\mathfrak{F}(S;\alpha)|^n d\alpha = 0 \tag{167}$$

and by the ergodic theorem, except for a set of values of α of zero measure, as r tends to 0 through a denumerable set of values,

$$\lim_{r\to 0}\frac{1}{V(r)}\int_R |\,\mathfrak{F}_{x_1,\ldots,x_n}(r;S;\alpha) - \mathfrak{F}_{x_1,\ldots,x_n}(S;\alpha)|^n dx_1\cdots dx_n = 0. \tag{168}$$

With this result as an aid, enabling us to show that the distribution of $\mathfrak{F}(S;\alpha)$ is only slightly affected by averaging within a small sphere with a given radius, or even within any small region near enough to a small sphere with a given radius, we may proceed as in (161) and (162) and form the chaos

$$\mathfrak{F}_k(S;\alpha)=\frac{1}{\mathfrak{N}(\mathfrak{P};\alpha)(2\pi k)^{n/2}}\int_\infty \mathfrak{F}_{x_1,\ldots,x_n}(S;\beta) \times(\mathfrak{M}(x_1,\cdots,x_n;\alpha)+\delta)\exp\frac{\left(-\sum_1^n x_j^2\right)}{2k}\,dx_1\cdots dx_n. \tag{169}$$

For almost all β, in each of the large cells of this chaos, (169) will have as nearly as we wish the same distribution as some $\mathfrak{F}_{x_1,\ldots,x_n}(S;\alpha)$, where $(x_1,\cdots,x_n)$ lies in the interior of the cell, remote from the boundary. These cells may so be determined that except for those filling an arbitrarily small proportion of space, all are convex regions with a minimum dimension greater than some given quantity.

To establish (143), it only remains to show that the average of a quantity depending on a chaos over a large cell tends to the same limit as its average over a large sphere. To show this, we only need to duplicate the argument of paragraph 4, where we prove the multidimensional ergodic theorem, for large pyramids with the origin as a corner, instead of for large spheres about the origin. We may take the shapes and orientations of these pyramids to form a denumerable assemblage, from which we may pick a finite assemblage which will allow us to approach as closely as we want to any cell for which the ratio of the maximum to the minimum distance from the origin within it does not exceed a given amount. It is possible to show that by discarding cells whose measure is an arbitrarily small fraction of the measure of all space, the remaining cells will have this property.

13. The physical problem. The transformation of a chaos. The statistical theory of a homogeneous medium, such as a gas or liquid, or a field of turbulence, deals with the problem, given the statistical configuration and velocity distribution of the medium at a given initial time, and the dynamical laws to which it is subject, to determine the configuration at any future time with respect to its statistical parameters. This of course is not a problem in the first instance of the history of the individual system, but of the entire ensemble, although in proper cases it is possible to show that almost all systems of the ensemble do actually share the same history, as far as certain specified statistical parameters are concerned.

The dynamical transformations of a homogeneous system have the very important properties, that they are independent of any choice of origin in time or in space. Leaving the time variable out of it, for the moment, the simplest space transformations of a homogeneous chaos $\mathfrak{F}(S;\alpha)$ which have this property are the polynomial transformations which turn it into

$$
\begin{aligned}
&K_0+\int K_1(x_1-y_1,x_2-y_2,\cdots,x_n-y_n)\,\mathfrak{F}_{y_1,\ldots,y_n}(S;\alpha)\,dy_1\cdots dy_n\\
(170)\qquad &+\cdots\\
&+\int\cdots\int K_\nu(x_1-y_1^{(1)},\cdots,x_n-y_n^{(1)},\cdots,x_1-y_1^{(\nu)},\cdots,\\
&\quad x_n-y_n^{(\nu)})\,\mathfrak{F}_{y_1^{(1)},\ldots,y_n^{(1)}}(S;\alpha)\cdots\\
&\qquad\qquad\mathfrak{F}_{y_1^{(\nu)},\ldots,y_n^{(\nu)}}(S;\alpha)\,dy_1^{(1)}\cdots dy_n^{(\nu)}.
\end{aligned}
$$

These are a sub-class of the general class of polynomial transformations

$$\begin{aligned} &K_0 + \int K_1(x_1, \cdots, x_n; y_1, \cdots, y_n) \mathfrak{F}_{y_1, \ldots, y_n}(S; \alpha)\, dy_1 \cdots dy_n \\ &\quad + \cdots \\ &\quad + \int \cdots \int K_\nu(x_1, \cdots, x_n; y_1^{(1)}, \cdots, y_n^{(1)}; \cdots; y_1^{(\nu)}, \cdots, y_n^{(\nu)}) \\ &\qquad \times \mathfrak{F}_{y_1(1), \ldots, y_n(1)}(S; \alpha) \cdots \mathfrak{F}_{y_1(\nu), \ldots, y_n(\nu)}(S; \alpha)\, dy_1^{(1)} \cdots dy_n^{(\nu)}. \end{aligned} \tag{171}$$

If a transformation of type (171) is invariant with respect to position in space, it must belong to class (170). On the other hand, in space of a finite number of dimensions and in any of the ordinary spaces of an infinite number of dimensions, polynomials are a closed set of functions, and hence every transformation may be approximated by a transformation of type (171).

A polynomial transformation such as (170) of a polynomial chaos yields a polynomial chaos. If then we can approximate to the state of a dynamical system at time 0 by a polynomial chaos, and approximate to the transformation which yields its status at time t by a polynomial transformation, we shall obtain for its state at time t, the approximation of another polynomial chaos. The theory of approximation developed in the last section will enable us to show this.

On the other hand, the transformation of a dynamical system induced by its own development is infinitely subdivisible in the time, and except in the case of linear transformations, this is not a property of polynomial transformations. Furthermore, when these transformations are non-linear, they are quite commonly not infinitely continuable in time. For example, let us consider the differential equation

$$\frac{\partial u}{\partial t} + u \frac{\partial u}{\partial x} = 0. \tag{172}$$

This corresponds to the history of a space-distribution of velocity transferred by particles moving with that velocity. Its solutions are determined by the equation

$$u(x, t) = u(x - tu(x, t), 0), \tag{173}$$

or if ψ is the inverse function of $u(x, 0)$,

$$x - tu(x, t) = \psi(u(x, t)). \tag{174}$$

Manifestly, if two particles with different velocities are allowed to move long enough to allow their space-time paths to cross, $u(x, t)$ will cease to exist as a single-valued function. This will always be the case for *some* value of t if $u(x, 0)$ is not constant, and for almost all values of t and α if it is a polynomial chaos.

By Lagrange's formula, (174) may be inverted into

$$u(x,t)=\sum_{0}^{\infty}\frac{(-t)^n}{n!}\frac{\partial^n}{\partial x^n}\{(u(x,0))^n\}. \tag{175}$$

In a somewhat generalized sense, the partial sums of this formally represent polynomial transformations of the initial conditions. However, it is only for a very special sort of bounded initial function, and for a finite value of the time, that they converge. It is only in this restricted sense that the polynomial transformation represents a true approximation to that given by the differential equation.

It will be seen that the useful application of the theory of chaos to the study of particular dynamical chaoses involves a very careful study of the existence theories of the particular problems. In many cases, such as that of turbulence, the demands of chaos theory go considerably beyond the best knowledge of the present day. The difficulty is often both mathematical and physical. The mathematical theory may lead inevitably to a catastrophe beyond which there is no continuation, either because it is not the adequate presentation of the physical facts; or because after the catastrophe the physical system continues to develop in a manner not adequately provided for in a mathematical formulation which is adequate up to the occurrence of the catastrophe; or lastly, because the catastrophe does really occur physically, and the system really has no subsequent history. The hydrodynamical investigations required in the case of turbulence are directly in the spirit of the work of Oseen and Leray, but must be carried much further.

The study of the history of a mechanical chaos will then proceed as follows: we first determine the transformation of the initial conditions generated by the dynamics of the ensemble. We then determine under what assumptions the initial conditions admit of this transformation for either a finite or an infinite interval of time. Then we approximate to the transformation for a given range of values of the time by a polynomial transformation. Then, having regard to a definition of distance between two functions determined by the transformation, we approximate to the initial chaos by a polynomial chaos. Next we apply the polynomial transformation to the polynomial chaos, and obtain an approximating polynomial chaos at time t. Finally, we apply our algorithm of the pure chaos to determine the averages of the statistical parameters of this chaos, and express these as functions of the time.

The results of such an investigation belong to a little-studied branch of statistical mechanics: the statistical mechanics of systems not in equilibrium. To study the classical, equilibrium theory of statistical mechanics by the methods of chaos theory is not easy. As yet we lack a method of representing all forms of homogeneous chaos, which will tell us by inspection when two differ merely by an equimeasure transformation of the parameter of distribution. In certain cases, in which the equilibrium is stable, the study of the history of a system with an arbitrary initial chaos will yield us for large values of t an approximation to equilibrium, but this will often fail to be so, particularly in the case of differentiable chaoses, or the only equilibrium may be that in which the chaos reduces to a constant.

MASSACHUSETTS INSTITUTE OF TECHNOLOGY.

ON HILL'S PERIODIC LUNAR ORBIT.

By Aurel Wintner.

Hill's desideratum. This paper deals with a question concerning the mathematical foundations of Hill's lunar theory, a theory developed in the first volume of this Journal [1] and considered by Poincaré as representing a turning point in the history of celestial mechanics.

From the mathematical point of view, the most original, but also the boldest, step in Hill's lunar theory consists of abandoning the classical method of variation of constants, i. e., Lagrange's theory of perturbations, and replacing it by successive applications of the method of infinitely many variables. Among mathematicians, Hill's name is remembered mainly for his theory of the lunar perigee, a theory depending on a linear Eigenwertproblem of specific type and leading to an infinite system of homogeneous linear equations. Hill's formal treatment of this problem by means of transcendental entire functions which are defined by infinite determinants has already been justified by Poincaré and has lead, via the investigations of Fredholm, to Hilbert's theory of linear transformations. Thus it is understandable that the introduction of the theory of infinite linear systems is generally considered as Hill's main contribution to pure analysis.

Actually, Hill's fundamental paper (*loc. cit.*[1]) does by no means deal with the linear Eigenwertproblem which leads to infinite determinants, but rather with a non-linear problem which leads to an infinite non-linear and non-recursive system of equations for infinitely many unknown Fourier constants; a system which seems to involve steps essentially more delicate than those which are implied by the application of infinite determinants in the linear problem. Furthermore, the linear problem presupposes that the non-linear problem has already been solved. In fact, the linear problem merely concerns the infinitesimal isoenergetic normal displacements of those periodic solutions of the non-linear equations of motion which lead to the infinite system of non-linear conditions.

---

[1] G. W. Hill, "Researches in the lunar theory," *American Journal of Mathematics*, vol. 1 (1878), pp. 5-26, 129-147, 245-260.

For a detailed presentation of Hill's programme in lunar theory, cf. pp. 195-236 (Chap. XI) of E. W. Brown's Treatise on the Lunar Theory (Cambridge, 1896), where further references may be found.

Hill was well aware of the novelty of the mathematical problem involved in his infinite non-linear implicit system of equations: "I regret that, on account of the difficulty of the subject . . . it does not appear that anything in the writings of Cauchy will help us to the conditions of convergence" (*loc. cit.*[1], p. 8). On the other hand, not only were Hill's numerical results in perfect agreement with the observed motion of the Moon, but also the terms of his expansions, obtained from the infinite non-linear system of equations in a formal and approximative manner, have shown, in contrast with the corresponding expansions of the other lunar theories, a surprisingly rapid decrease in absolute value; so that Hill (*loc. cit.*[1], p. 8) describes the probable situation as follows: "There cannot be a reasonable doubt that, in all cases, where we are compelled to employ infinite series in the solution of a problem, analysis is capable of being perfected to the point of showing us within what limits our solution is legitimate, and also giving us a limit which its error cannot surpass."

The analytical problem. The problem implied by this desideratum of Hill was treated by the writer several years ago.[2] The method consisted in proving (*loc. cit.*[2], pp. 263-265) a general existence theorem on infinite non-linear implicit systems which turned out to be the appropriate tool for dealing with Hill's problem and, at the same time, gave explicit estimates of the error terms which, in Hill's procedure, result by stopping the process of the comparison of coefficients at a certain stage.

The object of the present paper is to simplify this direct existence and convergence proof in such a way as to make it adaptable to sharper estimates. For this end, it will be convenient to use an existence theorem on infinite non-linear implicit systems which is somewhat more general[3] than the one used *loc. cit.*[2], and takes into account the peculiar situation in orders of magnitude which is an essential point in the problem. The proof of this appropriate existence theorem hardly differs from the proof of the more particular existence theorem which was used *loc. cit.*[2] and will, therefore, be omitted. Nevertheless, the paper is readable in itself.

Due to this general existence theorem, the treatment of Hill's problem becomes free of any ad hoc device, proceeds exactly along the lines which were followed by Hill in a formal way, and, in addition, takes into account the successive decrease of the orders of magnitude; thus proving, and also explaining, the convergence of Hill's process and supplying the realization of Hill's desideratum.

---

[2] A. Wintner, "Zur Hill'schen Theorie der Variation des Mondes," *Mathematische Zeitschrift*, vol. 24 (1925), pp. 259-265.

[3] Cf. A. Wintner, "Über die Differentialgleichungen der Himmelsmechanik," *Mathematische Annalen*, vol. 96 (1926), pp. 284-312.

It should be mentioned that the question as to the existence of Hill's family of periodic lunar orbits is implied by, but is not identical with, the question as to the legitimacy of Hill's bold method of infinitely many variables In fact, the question as to the existence of these periodic orbits can be settled according to Birkhoff, in a rather simple and intuitive manner.[4]

The equations of condition. Let x, y be the geocentric coördinates of the Moon in the rotating (i. e., synodic) Cartesian plane of Hill's restricted problem of three bodies. Then, if t denotes the time and m ($\gtrless 0$) an integration constant which varies in the neighborhood of $m = \pm 0$, Hill's aim (*loc. cit.*[2], pp. 129-131) is to find a continuous family of periodic solutions of the form

$$(1) \qquad x = \sum_{i=-\infty}^{+\infty} a_i \cos(2i+1)t/m, \qquad y = \sum_{i=-\infty}^{+\infty} a_i \sin(2i+1)t/m$$

where the Fourier coefficients a_i are unknown and are to be determined as real analytic functions,

$$(2) \qquad a_i = a_i(m) \qquad (i = 0, \pm 1, \pm 2, \cdots)$$

of the integration constant m, which is the parameter of the family (1).

One can replace m by the varying period

$$(3) \qquad T = 2\pi m \qquad (m \gtrless 0)$$

of (1) and also by Jacobi's constant $C = C(m)$, i. e., by the (relative) energy, where, however, $C(m) \rightarrow +\infty$ whether $m \rightarrow +0$ or $m \rightarrow -0$, i. e. whether the solution belonging to a small m is direct or retrograde. The numerical value, m_0, of m in the case of the Moon is, of course, positive and somewhat less than $1/12$:

$$(4) \qquad m_0 = 0.08084 \cdots$$

(cf. *loc. cit.*[1], p. 146).

Now, Hill's method of calculating the unknown family (1) of periodic solutions, i. e., the infinitely many unknown functions (2), consists simply of a direct substitution of (1) into the differential equations of motion and their Jacobi integral. This gives, by comparison of the coefficients in the resulting Fourier series, infinitely many implicit conditions for the infinitely many functions (2). Since the equations of motion are non-linear, so are these implicit conditions. In fact, Hill finds after several ingenious transformations that the infinite system of conditions which are to be satisfied by the functions (2) consists, on the one hand, of the equation[5]

---

[4] G. D. Birkhoff, "The restricted problem of three bodies," *Rendiconti del Circolo Matematico di Palermo*, vol. 39 (1915), Part III, § 12. Cf. also A. Wintner, Sortengenealogie, Hekubakomplex und Gruppenfortsetzung," *Mathematische Zeitschrift*, vol. 34 (1931), § 5 (Fig. 1) and § 7. A Russian paper by A. Liapounoff is reviewed in the *Fortschritte der Mathematik*, vol. 26 (1895), p. 1103.

$$(5)\qquad \{\sum_{i=-\infty}^{+\infty} a_i\}^2 \sum_{i=-\infty}^{+\infty} \{4i^2 + 4i + 1 + \overline{4i + \tfrac{1}{2}m} + 3m^2\}a_i = m^2$$

and, on the other hand, of the simultaneous infinite system[6]

$$(6)\qquad \sum_{i=-\infty}^{+\infty} \{[j,i]a_i a_{i-j} + m^2[j]a_i a_{-i+j-1} + m^2(j)a_i a_{-i-j-1}\} = 0;$$
$$(j = \pm 1, \pm 2, \cdots),$$

where $[j,i]$, $[j]$, (j) are rational functions of the independent parameter m, namely[7]

$$(7)\qquad [j,i] = -\frac{i}{j}\,\frac{4(j-1)i + 4j^2 + 4j - 2 - 4(i-j+1)m + m^2}{2(4j^2-1) - 4m + m^2},$$

$$(8)\qquad [j] = -\frac{3}{16j^2}\,\frac{4j^2 - 8j - 2 - 4(j+2)m - 9m^2}{2(4j^2-1) - 4m + m^2},$$

$$(9)\qquad (j) = -\frac{3}{16j^2}\,\frac{20j^2 - 16j + 2 - 4(5j-2)m + 9m^2}{2(4j^2-1) - 4m + m^2},$$

with the understanding that $j = \pm 1, \pm 2, \cdots$ but $i = 0, \pm 1, \pm 2, \cdots$.

Accordingly, the problem requires a proof to the effect that the infinite system (6), (5) of non-linear and non-recursive equations defines the infinitely many unknown functions (2), and that, for a fixed value (3) under consideration (and, in particular, for the fixed value (4) of m) the series (1) converge to functions $x = x(t)$, $y = y(t)$ which have second derivatives. For then (1) is a solution of the equations of motion, since (6), (5) represent all of the conditions obtained by the method of undetermined Fourier coefficients. Needless to say, $x = x(t)$, $y = y(t)$ must turn out to be analytic functions of t, the equations of motion being analytic. Incidentally, the functions (2) of m also prove to be analytic.

---

[5] G. W. Hill, *loc. cit.*[1], p. 145. This condition of Hill, when written in the above form (5), has as its right-hand member the number κ, instead of m^2. Actually, $\kappa = m^2$, if one chooses units in the usual way. In fact, Hill's n' denotes the constant angular velocity of the synodic coördinate system, while his μ is the total mass (*loc. cit.*[1], p. 9). Furthermore, $m = n'/(n - n')$ and $\kappa = \mu/(n - n')^2$ is his notation (*loc. cit.*[1], p. 9). Hence $\kappa = m^2$, if one chooses, as usual, $n' = 1$ and $\mu = 1$.

[6] *Ibid.*, p. 135. For reasons which will become apparent in (16) below, Hill's $[j]/m^2$ and $(j)/m^2$ have been denoted by $[j]$ and (j) respectively, while Hill's $[j,i]$ is the same as the above $[j,i]$.

[7] *Ibid.*, pp. 134-135. The factor m^2 of Hill's $[j]$ and (j) has been dropped in accordance with the preceding footnote.

Since the given rational functions (7), (8), (9) of the independent parameter m do not contain the unknown functions (2), the portion (6) of the complete system of conditions (5), (6) is homogeneous in the $a_i = a_i(m)$ and can, therefore, only determine the ratios

$$(10)\quad b_i = a_i : a_0 \equiv a_i(m) : a_0(m) = b_i(m) \qquad (i = \pm 1, \pm 2, \cdots)$$

of the unknown functions (2). If the functions (10) are known, the functions (2) follow from the remaining condition (5). In fact, the expression on the left of (5) is an infinite trilinear form in the $a_i = a_i(m)$; so that, if the functions (10) of m are known, a_0 follows as a function of m by the extraction of a cubic root. (Actually, the exponent of the radicand in m turns out to be $\frac{2}{3}$ instead of being merely $\frac{1}{3}$; and this $\frac{2}{3}$ expresses the infinitesimal validity of Kepler's third law).

Accordingly, the infinite system (5), (6) for the unknown functions (2) can be replaced by the infinite system (6) for the unknown functions (10).

A prepared form of the equations of condition. It is seen from (7) that

$$(11)\qquad [j,j] = -1 \quad\text{and}\quad [j,0] = 0 \qquad (j = \pm 1, \pm 2, \cdots)$$

for every m. Hence, (6) can be written as

$$(12)\quad a_0 a_j = \sum_{i=-\infty}^{+\infty}{}' [j,i] a_i a_{i-j} + m^2 [j] \sum_{i=-\infty}^{+\infty}{}'' a_i a_{-i+j-1}$$
$$+ m^2 (j) \sum_{i=-\infty}^{+\infty}{}''' a_i a_{-i-j-1} + 2m^2 [j] a_0 a_{j-1} + 2m^2 (j) a_0 a_{-j-1},$$
$$(j = \pm 1, \pm 2, \cdots),$$

where, corresponding to (11), the prime of the first sum means that

$$(13_1)\qquad i \neq j \text{ and } i \neq 0 \text{ in } \sum_{i=-\infty}^{+\infty}{}', \qquad (j = \pm 1, \pm 2, \cdots),$$

while two primes denote that

$$(13_2)\qquad i \neq j-1 \text{ and } i \neq 0 \text{ in } \sum_{i=-\infty}^{+\infty}{}'', \qquad (j = \pm 1, \pm 2, \cdots),$$

finally three primes that

$$(13_3)\qquad i \neq -j-1 \text{ and } i \neq 0 \text{ in } \sum_{i=-\infty}^{+\infty}{}''', \qquad (j = \pm 1, \pm 2, \cdots).$$

Since neither of the indices k, l of an arbitrary term $\cdots a_k a_l$ of the three sums (13_1)-(13_3) occurring in (12) vanishes, the homogeneous equations (12) for the unknown ratios (10) appear in the form

$$(14)\quad b_j = \sum_{i=-\infty}^{+\infty}{}' [j,i] b_i b_{i-j} + m^2 [j] \sum_{i=-\infty}^{+\infty}{}'' b_i b_{-i+j-1} + m^2 (j) \sum_{i=-\infty}^{+\infty}{}''' b_i b_{-i-j-1} + 2m^2 [j] b_{j-1} + 2m^2 (j) b_{-j-1}, \quad (j = \pm 1, \pm 2, \cdots),$$

if one divides (12) by a_0^2.

Now consider, instead of the ratios (10), the unknowns c_j which are defined by

$$(15)\quad b_j(m) \equiv b_j = m^{|j|+1} c_j, \text{ so that } c_j = c_j(m) \qquad (j = \pm 1, \pm 2, \cdots).$$

Then the j-th of the equations (14) appears, upon division by $m^{|j|+1}$, in the form

$$(16)\quad c_j = m\{ \sum_{i=-\infty}^{+\infty}{}' [j,i] m^{\alpha_{ij}} c_i c_{i-j} + [j] \sum_{i=-\infty}^{+\infty}{}'' m^{\beta_{ij}} c_i c_{-i+j-1} + (j) \sum_{i=-\infty}^{+\infty}{}''' m^{\gamma_{ij}} c_i c_{-i-j-1} + 2[j] m^{\delta_j} c_{j-1} + 2(j) m^{\epsilon_j} c_{-j-1} \}, \quad (j = \pm 1, \pm 2, \cdots),$$

where the exponents are integers; in fact, corresponding to (13_1)-(13_3),

$$(17_1)\qquad \alpha_{ij} = |i-j| + |i| - |j| \text{ in } \sum_{i=-\infty}^{+\infty}{}' ;$$

$$(17_2)\qquad \beta_{ij} = |i+1-j| + |i| - |j| + 2 \text{ in } \sum_{i=-\infty}^{+\infty}{}'' ;$$

$$(17_3)\qquad \gamma_{ij} = |i+1+j| + |i| - |j| + 2 \text{ in } \sum_{i=-\infty}^{+\infty}{}''' ;$$

while obviously,

$$(18_1)\quad \delta_j = |j-1| - |j| + 1; \qquad (18_2)\quad \epsilon_j = |j+1| - |j| + 1.$$

Since $j = \pm 1, \pm 2, \cdots$, comparison of (17_ν) with (13_ν), where $\nu = 1, 2, 3$, shows that

(19) the integers $\alpha_{ij}, \beta_{ij}, \gamma_{ij}, \delta_j, \epsilon_j$ are non-negative,

the inequalities $\delta_j \geqq 0$ and $\epsilon_j \geqq 0$ being obvious from (18_1)-(18_2).

Estimates for the coefficient functions. In what follows, use will be made of the notation

$$(20)\qquad f^* \equiv f^*(m) = \sum_{n=0}^{\infty} |c_n| m^n, \text{ if } f \equiv f(m) = \sum_{n=0}^{\infty} c_n m^n$$

is a regular analytic function of the independent variable m in the vicinity of the origin $m = 0$.

Since $j \neq 0$, any of the infinitely many rational functions (7), (8), (9) is regular at $m = 0$; so that (20) defines $[j,i]^*$, $[j]^*$, $(j)^*$ as power series in m with real non-negative coefficients. Now, all these power series are convergent in the circle $|m| \leqq 1$ of the complex m-plane. Furthermore, there exists a bound C which is independent of j and i and is such that

$$(21) \qquad |[j,i]^*| < C\,|i/j|^2 + C\,|i/j| \text{ for } |m| \leqq 1$$

and

$$(22) \qquad |[j]^*| < C/j^2 \text{ and } |(j)^*| < C/j^2 \text{ for } |m| \leqq 1.$$

In fact, if $f(m)$ is regular analytic in a circle $|m| < R$, and if $|f(m)| < M$ in this circle, then, on using the notations (20), one has $|c_n| < M/R^n$ (Cauchy), and so, if $R > 1$,

$$|f^*(m)| < MR/(R-1) \text{ for } |m| \leqq 1.$$

Consequently, (21) is true if there exists two constants M and R such that $R > 1$ and all of the functions (7) are regular and less than M in absolute value in the circle $|m| < R$. Now, on dividing the numerator and denominator of (7) by $2(4j^2-1)$, one sees that M and $R > 1$ exist if the rational functions

$$(23) \qquad \left\{ 1 - \frac{4m - m^2}{2(4j^2-1)} \right\}^{-1}, \text{ where } j = \pm 1, \pm 2, \cdots,$$

are regular and uniformly bounded in a circle $|m| < R$, where $R > 1$. Finally, this condition is satisfied by $R = 8/7$, since if $|m| < 1 + 1/7$, then

$$|4m - m^2| < (1 + 1/7)(4 + 1 + 1/7) < 6 \leqq 2(4j^2 - 1) \text{ for } j = \pm 1, \pm 2, \cdots.$$

This proves (21); while (22) follows in the same way from (8), (9) as (21) did from (7).

A consequence of (21) and (22) is that if the variables $m;\ c_1, c_{-1}, c_2, c_{-2}, \cdots$ are restricted by the inequalities

$$(24) \qquad |m| \leqq 1; \quad |c_j| \leqq 1/j^2 \qquad (j = \pm 1, \pm 2 \cdots)$$

but are otherwise arbitrary, then the estimates

$$(25) \qquad \sum_{i=-\infty}^{+\infty}{}' \left| [j,i]^* \, m^{\alpha_{ij}} c_i c_{i-j} \right| < B/j^2, \qquad (j = \pm 1, \pm 2 \cdots)$$

and

$$(26) \qquad \begin{aligned} |[j]^*| \sum_{i=-\infty}^{+\infty}{}'' \left| m^{\beta_{ij}} c_i c_{-i+j-1} \right| &< B/j^2, \\ |(j)^*| \sum_{i=-\infty}^{+\infty}{}''' \left| m^{\gamma_{ij}} c_i c_{-i-j-1} \right| &< B/j^2, \end{aligned} \qquad (j = \pm 1, \pm 2 \cdots)$$

hold for a sufficiently large constant B which, in addition, can be chosen such that

$$(27) \qquad \begin{aligned} 2\,|[j]^* \, m^{\delta_j} \, c_{j-1}| &< B/j^2, \\ 2\,|(j)^* \, m^{\epsilon_j} \, c_{-j-1}| &< B/j^2, \end{aligned} \qquad (j = \pm 1, \pm 2 \cdots).$$

In fact, it is clear from (19) and (21) that, in the domain (24), the expression on the left of (25) is less than the product of the constant C and of

$$\sum_{i=-\infty}^{+\infty}{}' i^2 j^{-2}\cdot 1\cdot i^{-2}(i-j)^{-2} + \sum_{i=-\infty}^{+\infty}{}' | ij^{-1} | \cdot 1\cdot i^{-2}(i-j)^{-2}.$$

Since the first of these sums can be written as

$$\sum_{i=-\infty}^{+\infty}{}' j^{-2}(i-j)^{-2} = j^{-2}\cdot\text{const.},$$

and the second as

$$\sum_{i=-\infty}^{+\infty}{}' | ji |^{-1}(i-j)^{-2} < | j |^{-1}\sum_{i=-\infty}^{+\infty}{}' (i-j)^{-2} < | j |^{-1}\cdot\text{const.}\sum_{i=0}^{+\infty}(i+| j |)^{-2}$$
$$= | j |^{-1}\cdot\text{const.}\sum_{i=|j|}^{+\infty} i^{-2} < | j |^{-1}\cdot\text{Const.}\, | j |^{-1} = \text{Const.}/j^2,$$

the proof of (25) is complete.

Furthermore, (19) and (22) show that, in the domain (24), the expression on the left of the first of the inequalities (26) is less than

$$Cj^{-2}\sum_{i=-\infty}^{+\infty}{}'' i^{-2}(-i+j-1)^{-2} < 2Cj^{-2}\sum_{i=1}^{+\infty} i^{-2} = \text{const.}/j^2.$$

This proves the first of the inequalities (26), while the second follows in the same way.

Finally, (27) is clear from (19), (22) and (24).

The existence theorem. Without regard to questions of convergence, let

$$\phi_n \equiv \phi_n(z_0, z_1, \cdots) = \Sigma\cdots\Sigma\, a_{i_1\ldots i_n} z_{i_1}\cdots z_{i_n},$$

where the n non-negative integers $i_1,\cdots,i_n$ need not be distinct, denote a form of degree n in the infinitely many variables $z_0, z_1,\cdots$, and suppose that the terms of this n-fold sum have been contracted completely, i. e. that the n equalities $i_1 = j_1,\cdots, i_n = j_n$ always imply the vanishing of at least one of the two constant coefficients $a_{i_1\ldots i_n}$, $a_{j_1\ldots j_n}$. On this assumption, associate with ϕ_n the unique form

$$\phi^*_n \equiv \phi^*_n(z_0, z_1, \cdots) = \Sigma\cdots\Sigma\, | a_{i_1\ldots i_n} | z_{i_1}\cdots z_{i_n}$$

(so that $\phi_n \equiv \phi^*_n$ if and only if all $a \geqq 0$). And let f^* denote the unique power series

$$f^* \equiv f^*(z_0, z_1, \cdots) = \sum_{n=0}^{\infty}\phi^*_n, \quad\text{if}\quad f \equiv f(z_0, z_1, \cdots) = \sum_{n=0}^{\infty}\phi_n$$

is a power series in the infinitely many variables $z_0, z_1,\cdots$; so that, corresponding to (20),

$$f^* = \sum_{n=0}^{\infty}\sum\sum\cdots\sum |a_{i_1\ldots i_n}| z_{i_1}\cdots z_{i_n} \text{ if } f = \sum_{n=0}^{\infty}\sum\sum\cdots\sum a_{i_1\ldots i_n} z_{i_1}\cdots z_{i_n}.$$

Let there be given an infinite sequence

(28) $$f_1(x; y_1, y_2, \cdots), \cdots, \quad f_k(x; y_1, y_2, \cdots), \cdots$$

of such power series $f(z_0, z_1, z_2, \cdots)$, where $z_0 = x$; $z_1 = y_1$, $z_2 = y_2, \cdots$. Suppose that there exist two positive constants and two sequences of positive constants, say

(29) $$\alpha; \mu \text{ and } B_1, \cdots, B_k, \cdots; M_1, \cdots, M_k, \cdots$$

such that, on the one hand,

(30) $$B_k/M_k \geqq \mu > 0, \qquad (k = 1, 2, \cdots)$$

and, on the other hand, the sequence of power series (28) satisfies the conditions

(31) $$f^*_k(\alpha; B_1, B_2, \cdots) \leqq M_k, \qquad (k = 1, 2, \cdots)$$

(which, in particular, imply that each of the power series (28) is absolutely convergent in the domain

(32) $$|x| \leqq \alpha; \quad |y_1| \leqq B_1, \cdots, \quad |y_k| \leqq B_k, \cdots$$

of infinitely many dimensions).

Now (cf. *loc. cit.*[2] and *loc. cit.*[3]), if the power series (28) in infinitely many variables are such that the conditions (30), (31) are satisfied by positive constants (29), then the infinite implicit system of equations

(33) $$y_k = x f_k(x; y_1, y_2, \cdots), \qquad (k = 1, 2, \cdots)$$

has in a sufficiently small fixed circle about the origin of the complex x-plane, namely at least in the circle

(34) $$|x| < \min(\alpha, \mu)$$

which is independent of k, exactly one regular analytic solution $y_k = y_k(x)$. Furthermore,

(35) $$|y_k(x)| < B_k \text{ for } |x| < \min(\alpha, \mu), \qquad (k = 1, 2, \cdots).$$

Finally, every $y_k(x)$ is real for real x if the coefficients of the power series (28) in infinitely many variables are real.

The proof of this theorem (*loc. cit.*[2] and *loc. cit.*[3]) is such as to supply not only the existence of the regular functions $y_k(x)$ and their estimates (35) but also convenient appraisals of the errors which are committed if one only carries out the process of the method of undetermined coefficients to a given stage.

Application of the existence theorem. Now, this existence theorem is applicable to the problem at hand, if one identifies

$$m;\ c_1, c_{-1}, c_2, c_{-2}, \cdots \text{ with } x;\ y_1, y_2, y_3, y_4, \cdots$$

respectively. In fact, the rational functions (7), (8), (9) of m are regular at $m = 0$ and can, therefore, be developed according to non-negative powers of $m = x$. Hence it is clear from (19) that the system (16) for the unknown functions $c_j = c_j(m)$ can be written in the form

$$(36) \qquad c_j = mF_j(m; c_1, c_{-1}, c_2, c_{-2}, \cdots) \qquad (j = \pm 1, \pm 2, \cdots)$$

where the F_j are given power series in the infinitely many variables m; c_j and have real coefficients. Furthermore, there exists a constant $A > 0$ such that

$$(37) \qquad F^*_j(1; 1^{-2}, 1^{-2}, 2^{-2}, 2^{-2}, \cdots) < Aj^{-2} \qquad (j = \pm 1, \pm 2, \cdots).$$

In fact, since (25), (26), (27) hold on the assumption (24), one can choose $A = 5B$, the function F_j in (36) being, by its definition, the function { } in (16).

On identifying (36) with (33), one sees from (37) that the conditions of the existence theorem, namely (31) and (30), are satisfied by

$$\alpha = 1;\ B_j = 1/j^2,\ M_j = A/j^2 \qquad (j = \pm 1, \pm 2, \cdots)$$

since (30) is then satisfied by $\mu = 1/A$. Thus, the radius of the circle (34), where $x = m$, becomes $\min(1, 1/A)$; while (35) states that $|c_j(m)| < 1/j^2$, hence $|c_j(m)| < 1$, in this circle. Furthermore, $c_j(0) = 0$, as seen by placing $m = 0$ in (36).

Since (36) is identical with (16), while (16) is identical with (14) in virtue of (15), it follows that the system (14) has in the domain

$$(38) \qquad |m| < \lambda, \text{ where } \lambda = \min(1, 1/A), \qquad (\beta < 1),$$

exactly one regular analytic solution (10), and that the Taylor series of the infinitely many functions $b_j(m)$ are of the form

$$(39) \qquad b_j(m) = m^{|j|+2} \sum_{n=0}^{\infty} \beta_{jn} m^n \qquad (j = \pm 1, \pm 2, \cdots),$$

where the constants β_{jn} are real, finally that the power series (29) are in their common domain (38) of convergence such that

$$(40) \qquad |b_j(m)| < |m|^{|j|}, \qquad (j = \pm 1, \pm 2, \cdots;\ |m| < \lambda < 1).$$

Now, (14) is, in virtue of (10), equivalent to (6), while (5) becomes a cubic condition for the single function $a_0 = a_0(m)$, since the ratios (10) are the known functions (39). Finally, direct substitutions readily show that the unique real analytic function $a_0(m)$ which is thus defined by (5) does not vanish for small $m \gtrless 0$.

This completes the treatment of Hill's desideratum concerning (1). In fact, $a_0 = a_0(m)$ being a known function of m, the coefficients (2) of (1) follow from (10) and (39) uniquely, while (40) assures that the series (2) are Fourier series of (analytic) periodic functions of t.

The numerical situation. It should be mentioned that, on assuming in a formal way that $b_j(m)$ is a power series, it easily follows[8] from (6) or (14) without any reference to an existence theorem regarding (33), that $b_j(m)$ must be divisible by $m^{|j|}$. Needless to say, this does not dispose of the necessity of an existence and convergence proof, although $|m| < 1$. In fact, the usual formal treatment[9] of (6) does not take into account the possibility that not only (40) might not hold on a circle (38) which is independent of j but also that some or all of the power series (39) might be divergent for every $m \neq 0$, although $b_j(m)$ is, in a formal way, divisible by $m^{|j|}$. Actually, it is clear a priori that if the periodic solution (1) exists at all for a fixed m which does not correspond to a collision path, then $|a_j| <$ const. $\vartheta^{|j|}$ for a suitable positive $\vartheta < 1$, since the periodic functions (1) must be analytic in the real variable t; so that (40) is by no means surprising, if the existence of (1) or (2) is assumed.

Loc. cit.[2], an existence theorem concerning (33) was applied to (14) instead of to (16), so that the numerical advantages of Hill's equations of condition were not completely utilized. Furthermore, the existence theorem applied *loc. cit.*[2] was a particular case of the one used above, since, instead of the general condition (30) used above, it was assumed that all B_k and all M_k are equal. This necessitated (*loc. cit.*[2]) the use of an absolute constant connected with a fact which is fundamental in Riemann's theory of trigonometrical series, namely with the fact that

$$\sum_{i=-\infty}^{+\infty}{}' \frac{1}{i^2(i-j)^2} = O\left(\frac{1}{j^2}\right) \text{ as } j \to \pm\infty.$$

All of this has made the numerical situation quite unfavorable, although it has allowed the proof of the fact[10] that the process is valid in a circle (38) whose radius is not less than 1/1[illegible] While this circle contains the value (4) belonging to the case of the Moon, Hill's numerical evidence indicates that the convergence must be much stronger than what one would expect of the closeness of the value (4) to the bound $1/12 = 0.08333\cdots$, if this bound were sharp. Now, for reasons explained before, the method of the present paper leads to bounds much better than $1/12$. The detailed numerical work to this effect is in progress.

---

[8] Cf. H. Poincaré, *Leçons de mécanique céleste*, vol. 2, part 2 (1909), pp. 35-36.

[9] Cf. H. Poincaré, *loc. cit.*[8], pp. 34-39.

[10] A. Wintner, "Über die Konvergenzfragen der Mondtheorie," *Mathematische Zeitschrift*, vol. 30 (1929), pp. 211-2[illegible].

Of course, a bound of this type, no matter how sharp, is necessarily unable to answer the important question as to the "natural termination" (Strömgren) of Hill's family of periodic orbits.[11] All that follows from the general theory is that such a "dynamical termination" exists.[12] As far as the problem of convergence in the vicinity of $m = \pm 0$ is concerned, it is clear that complex singularities in m, which have no significance for the problem of termination, can act unfavorably.

Correspondingly, it is, according to Hill (*loc. cit.*[1]), convenient to carry out the expansions, not according to the powers of m but rather according to the powers of a linear function $\tilde{m} = \tilde{m}(m)$ of m which vanishes with m and is chosen so as to transform those complex singularities in m which are formally indicated by (23) into positions which are not near the real axis. Clearly, this amounts to what is called, in the theory of divergent series, an Euler transformation. For instance, if m is positive and small, one improves on the closeness of the complex singularities of the rational functions (23) for $j = \pm 1, \pm 2, \cdots$, if one replaces m by the expansion parameter

$$m' = m/(1+m) \tag{41}$$

of Delaunay's lunar theory. It is easy to rewrite the existence proof of the present paper from m to this m' or to Hill's $\tilde{m} = \tilde{m}(m)$, and then to calculate the corresponding bounds, thus improved. This numerical work also is in progress.

Needless to say, a transformation of the type (41) is well adapted only to the case of direct paths, i. e., to the region ($m > 0$) of the Moon. In view of a result of Birkhoff,[13] the initial behavior of the retrograde paths ($m < 0$), which are not excluded in (38), is a much easier problem than the corresponding behavior of the direct paths.

THE JOHNS HOPKINS UNIVERSITY.

---

[11] Cf. A. Wintner, "Almost periodic functions and Hill's theory of lunar perigee," *American Journal of Mathematics*, vol. 59 (1937), pp. 795-802, footnote 12 on p. 798.

[12] A. Wintner, "Beweis des E. Strömgrenschen dynamischen Abschlussprincips," *Mathematische Zeitschrift*, vol. 34 (1931), pp. 321-349. Cf. G. D. Birkhoff, *Sur le problème restreint des trois corps*, Second mémoire, chap. III, § 5, *Annali della R. Scuola Normale Superiore di Pisa*, serie 2, vol. 5 (1936).

[13] G. D. Birkhoff, *loc. cit.*[4], § 18. Cf. also A. Wintner, *loc. cit.*[4], § 8–§ 9.